OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 16:21:08 ; Search time 6864.35 Seconds
(without alignments)
11877.277 Million cell updates/sec

Title: US-09-828-068-1
Perfect score: 3896
Sequence: 1 cgcggctgtcggagcaaacg.....gtcaacaccggagaatttac 3896

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3892.8	99.9	3896	8	AF326768	AF326768 Oryza sat
2	2952.2	75.8	150594	8	AP001859	AP001859 Oryza sat
c 3	62.8	1.6	7218	6	I66494	I66494 Sequence 14
4	48.2	1.2	141041	2	AC026135	AC026135 Homo sapi
c 5	48.2	1.2	141095	2	AC092925	AC092925 Homo sapi
c 6	47.2	1.2	189409	2	AL591122	AL591122 Homo sapi
c 7	46.2	1.2	151076	2	AC055845	AC055845 Homo sapi
8	45.4	1.2	145120	8	AP003434	AP003434 Oryza sat
c 9	45.2	1.2	1141	6	AX083744	AX083744 Sequence
10	45.2	1.2	93100	9	AL162614	AL162614 Human DNA
11	45.2	1.2	114980	2	AC087825	AC087825 Homo sapi
12	45.2	1.2	158392	2	AC022048	AC022048 Homo sapi
c 13	44.4	1.1	3400	14	SH1UL45A	L20708 Pseudorabie
c 14	44.2	1.1	65478	2	AC107868	AC107868 Mus muscu
15	44.2	1.1	95097	9	AL133232	AL133232 Human DNA
c 16	44	1.1	155204	2	AC007926	AC007926 Trypanoso
c 17	43.6	1.1	203220	2	AC097681	AC097681 Rattus no
18	43.4	1.1	149704	2	AC103138	AC103138 Rattus no
c 19	43.2	1.1	2642	3	DDIGBF	L29075 Dictyosteli
c 20	43	1.1	3360	2	AC014126	AC014126 Drosophil
21	43	1.1	201313	3	AC010688	AC010688 Drosophil
22	43	1.1	201978	2	AC079533	AC079533 Mus muscu
23	43	1.1	280887	3	AE003526	AE003526 Drosophil
c 24	42.8	1.1	125020	9	AF429315	AF429315 Homo sapi
25	42.6	1.1	696	8	CNS0192T	AL111116 Botrytis
26	42.6	1.1	179075	2	AC018518	AC018518 Homo sapi
27	42.6	1.1	213621	2	AC104376	AC104376 Homo sapi
c 28	42.6	1.1	227518	2	AC091287	AC091287 Mus muscu
29	42.6	1.1	256172	2	AC005139	AC005139 Plasmodiu
30	42.4	1.1	1747	8	NTHSP82	X63195 N.tabacum m
31	42.4	1.1	49376	2	AC023803	AC023803 Mus muscu
c 32	42.2	1.1	201882	2	AC103345	AC103345 Rattus no
33	42.2	1.1	313450	1	AL596170	AL596170 Listeria
c 34	42	1.1	275	8	AY023518	AY023518 Oryza sat
35	42	1.1	125020	9	AF429315	AF429315 Homo sapi
36	42	1.1	180615	2	AC053543	AC053543 Homo sapi
37	42	1.1	181139	2	AC090668	AC090668 Homo sapi
38	41.8	1.1	20752	3	CEC06A1	Z49886 Caenorhabdi
c 39	41.8	1.1	70213	2	AC103789	AC103789 Homo sapi
c 40	41.8	1.1	129245	9	AL359547	AL359547 Human DNA
c 41	41.8	1.1	183033	2	AC016272	AC016272 Homo sapi
42	41.8	1.1	199508	2	AC098555	AC098555 Rattus no
c 43	41.6	1.1	74589	8	AB010068	AB010068 Arabidops
c 44	41.6	1.1	253305	3	PFMAL3P7	AL034559 Plasmodiu
c 45	41.4	1.1	2848	1	AF269518	AF269518 Staphyloc

ALIGNMENTS

RESULT 1
AF326768
LOCUS AF326768 3896 bp mRNA linear PLN 05-SEP-2001
DEFINITION Oryza sativa embryonic flower 1-like protein mRNA, complete cds.
ACCESSION AF326768
VERSION AF326768.1 GI:15430698
KEYWORDS .
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 3896)
AUTHORS Aubert,D., Chen,L., Moon,Y.H., Martin,D., Castle,L.A., Yang,C.H. and Sung,Z.R.
TITLE Emf1, a novel protein involved in the control of shoot architecture and flowering in arabidopsis
JOURNAL Plant Cell 13 (8), 1865-1875 (2001)
MEDLINE 21380417
PUBMED 11487698
REFERENCE 2 (bases 1 to 3896)
AUTHORS Moon,Y.-H., Chen,L. and Sung,Z.R.
TITLE Direct Submission

```
  JOURNAL   Submitted (06-DEC-2000) Plant and Microbial Biol. Dept., University
            of California, Berkeley, 361 Koshland Hall, Berkeley, CA 94720, USA
FEATURES             Location/Qualifiers
     source          1. .3896
                     /organism="Oryza sativa"
                     /db_xref="taxon:4530"
     CDS             241. .3414
                     /note="EMF1"
                     /codon_start=1
                     /product="embryonic flower 1-like protein"
                     /protein_id="AAK98529.1"
                     /db_xref="GI:15430699"
                     /translation="MEIVAVDQEGARVVGTNCMLARGGTGAVAPVLELTATPRQDAAA
                     EAGVDEPAQHQCEHFSIRGYVALLQKKDPKFCSLSRIFHDQKKCDEHKASSSPFSVAK
                     FRRWDCSKCLDKLKTSDNGTAPRTLPAKQNGTSDGCSITFVRSTFVPASVGSQKVSPS
                     TQSSQGKNADRSTLPKSVQEGNDSKCNAPSGKNGAAEANTDSPMKDLQGPAQNYDVAA
                     NVSEDNTSVDVGALPEVPQITWHIEVNGADQPPSTPKLSEVVLKRNEDENGKTEETLV
                     AEQCNLTKDPNPMSGKERDQVAEQCNLTKDPKPVSGQKCEQICNEPCEEVVLKRSSKS
                     KRKTDKKLMKKQQHSKKRTAQADVSDAKLCRRKPKKVRLLSEIVNANQVEDSRSDEVH
                     RENAADPCEDDRSTIPVPMEVSMDIPVSNHTVGEDGLKSSKNKTKRKYSDVVDDGSSL
                     MNWLNGKKKRTGSVHHTVAHPAGNLSNKKVTPTASTQHDDENDTENGLDTNMHKTDVC
                     QHVSEISTQRCSSKGKTAGLSKGKTHSAASTKYGGESTRNGQNIHVLSAEDQCQMETE
                     NSVLSHSAKVSPAEHDIQIMSDLHEQSLPKKKKKQKLEVTREKQTMIDDIPMDIVELL
                     AKNQHERQLMTETDCSDINRIQSKTTADDDCVIVAAKDGSDYASSVFDTNSQQKSLAS
                     QSTQKELQGHLALTTQESPHPQNFQSTQEQQTHLRMEEMVTIAASSPLFSHHDDQYIA
                     EAPTEHWGRKDAKKLTWEQFKATTRNSPAATCGAQFRPGIQAVDLTSTHVMGSSSNYA
                     SRQPVIAPLDRYAERAVNQVHARNFPSTIATMEASKLCDRRNAGQVVLYPKESMPATH
                     LLRMMDPSTLASFPNYGTSSRNQMESQLHNSQYAHNQYKGSTSTSYGSNLNGKIPLTF
                     EDLSRHQLHDLHRPLRPHPRVGVLGSLLQKEIANWSENCGTQSGYKLGVSTGITSHQM
                     NRKEHFEALNSGMFSAKWNALQLGSVSSSAYFLSARNSIAQSWTRGKGKMVHPLDRFV
                     RQDICITNKNPADFTTISNDNEYMDYR"
  BASE COUNT     1175 a    791 c    935 g    995 t
  ORIGIN
```

```
  Query Match             99.9%;  Score 3892.8;  DB 8;  Length 3896;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 3894;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 cgcggctgtcggagcaaacgcaaacccccccaggttgttctagcgtgtgcagcggctagct 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CGCGGCTGTCGGAGCAAACGCAAACCCCCCCAGGTTGTTCTAGCGTGTGCAGCGGCTAGCT 60

Qy      61 gattgattgtcttctgtgatatatccagagctcgtgttttgtggtttgtggtttgtggtt 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GATTGATTGTCTTCTGTGATATATCCAGAGCTCGTGTTTTGTGGTTTGTGGTTTGTGGTT 120

Qy     121 tgtgcttggattgttgatgtgctaattcgcggcgttacaagatcactgctggattgatat 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TGTGCTTGGATTGTTGATGTGCTAATTCGCGGCGTTACAAGATCACTGCTGGATTGATAT 180

Qy     181 tgagttgtgcctcggctgtgctggctgtgtgttgattctctcctcgtcgtggtgatcgat 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TGAGTTGTGCCTCGGCTGTGCTGGCTGTGTGTTGATTCTCTCCTCGTCGTGGTGATCGAT 240

Qy     241 atggagattgttgcagtagatcaggagggagctcgtgttgttgggacgaactgtatgctt 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 ATGGAGATTGTTGCAGTAGATCAGGAGGGAGCTCGTGTTGTTGGGACGAACTGTATGCTT 300

Qy     301 gctcgtggtggaactggtgctgtagcgccagtgttggagctgacagcgacgcctcgtcag 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GCTCGTGGTGGAACTGGTGCTGTAGCGCCAGTGTTGGAGCTGACAGCGACGCCTCGTCAG 360

Qy     361 gatgcagccgctgaagctggtgtagacgaaccggcacaacaccaatgcgagcatttctcc 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GATGCAGCCGCTGAAGCTGGTGTAGACGAACCGGCACAACACCAATGCGAGCATTTCTCC 420

Qy     421 ataagagggtatgttgctcttcttcagaagaaggatccaaaattctgctctctatctcgg 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 ATAAGAGGGTATGTTGCTCTTCTTCAGAAGAAGGATCCAAAATTCTGCTCTCTATCTCGG 480

Qy     481 attttccatgaccagaaaaaatgtgatgaacacaaagctagttcaagcccatttttctgta 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 ATTTTCCATGACCAGAAAAAATGTGATGAACACAAAGCTAGTTCAAGCCCATTTTCTGTA 540

Qy     541 gcaaagtttcgacgatgggattgctcgaagtgcttggataagttgaaaacttcagataat 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 GCAAAGTTTCGACGATGGGATTGCTCGAAGTGCTTGGATAAGTTGAAAACTTCAGATAAT 600

Qy     601 ggaacagcaccaagaactcttcccgcaaagcagaatggcacaagtgatggttgctccatc 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 GGAACAGCACCAAGAACTCTTCCCGCAAAGCAGAATGGCACAAGTGATGGTTGCTCCATC 660

Qy     661 acatttgttcggagcacttttgtgcctgctagtgttggttcccaaaaagtgtctcctagc 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 ACATTTGTTCGGAGCACTTTTGTGCCTGCTAGTGTTGGTTCCCAAAAAGTGTCTCCTAGC 720

Qy     721 acacaatcatctcaagggaagaatgctgatagatcaactcttccaaagagtgtgcaagaa 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 ACACAATCATCTCAAGGGAAGAATGCTGATAGATCAACTCTTCCAAAGAGTGTGCAAGAA 780

Qy     781 ggcaatgactccaaatgcaatgcgccttctggcaagaatggagctgctgaggccaatact 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 GGCAATGACTCCAAATGCAATGCGCCTTCTGGCAAGAATGGAGCTGCTGAGGCCAATACT 840

Qy     841 gattcaccaatgaaagatttgcaagggccagcccaaaattatgatgtggcagcaaatgtc 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 GATTCACCAATGAAAGATTTGCAAGGGCCAGCCCAAAATTATGATGTGGCAGCAAATGTC 900

Qy     901 tctgaggacaacacttctgttgatgttggggctttacctgaagttccccagattacatgg 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 TCTGAGGACAACACTTCTGTTGATGTTGGGGCTTTACCTGAAGTTCCCCAGATTACATGG 960

Qy     961 cacatagaagtaaatggtgcagatcaacctccatccactccaaaacttttctgaagtggtc 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 CACATAGAAGTAAATGGTGCAGATCAACCTCCATCCACTCCAAAACTTTCTGAAGTGGTC 1020

Qy    1021 ctcaaaagaaatgaagatgaaaatggaaaaactgaagagactcttgttgctgagcagtgc 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 CTCAAAAGAAATGAAGATGAAAATGGAAAAACTGAAGAGACTCTTGTTGCTGAGCAGTGC 1080

Qy    1081 aatttgaccaaagatcctaacccaatgtctggaaaggaacgtgatcaggttgctgagcag 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 AATTTGACCAAAGATCCTAACCCAATGTCTGGAAAGGAACGTGATCAGGTTGCTGAGCAG 1140

Qy    1141 tgcaatttgaccaaagatccgaaaccagtgtctgggcagaaatgtgagcagatctgcaat 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 TGCAATTTGACCAAAGATCCGAAACCAGTGTCTGGGCAGAAATGTGAGCAGATCTGCAAT 1200

Qy    1201 gagccatgtgaagaagttgttctcaaaagaagctccaaatctaagaggaagacggataag 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 GAGCCATGTGAAGAAGTTGTTCTCAAAAGAAGCTCCAAATCTAAGAGGAAGACGGATAAG 1260

Qy    1261 aagttgatgaagaagcagcagcacagcaagaaacgcactgcccaggctgatgtttcagat 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 AAGTTGATGAAGAAGCAGCAGCACAGCAAGAAACGCACTGCCCAGGCTGATGTTTCAGAT 1320

Qy    1321 gcaaagctttgtcggagaaagccaaaaaaggtgcggcttctatcagaaattataaatgct 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 GCAAAGCTTTGTCGGAGAAAGCCAAAAAAGGTGCGGCTTCTATCAGAAATTGTAAATGCT 1380

Qy    1381 aaccaggttgaggattctagaagtgacgaagttcatcgtgaaaatgccgctgatccctgt 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 AACCAGGTTGAGGATTCTAGAAGTGACGAAGTTCATCGTGAAAATGCCGCTGATCCCTGT 1440

Qy    1441 gaggatgatagaagtaccatcccggtcccgatggaagtaagcatggatattcctgttagc 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 GAGGATGATAGAAGTACCATCCCGGTCCCGATGGAAGTAAGCATGGATATTCCTGTTAGC 1500

Qy    1501 aaccatacagtgggagaagatgggttaaaatcaagtaagaacaagacaaaacgcaaatac 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1501 AACCATACAGTGGGAGAAGATGGGTTAAAATCAAGTAAGAACAAGACAAAACGCAAATAC 1560

Qy    1561 tctgatgttgtagatgatggatcatcacttatgaactggctgaatggaaaaaagaaaaga 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1561 TCTGATGTTGTAGATGATGGATCATCACTTATGAACTGGCTGAATGGAAAAAAGAAAAGA 1620
```

```
Qy   1621 actggaagtgtgcatcacacagttgctcatccagctgggaatttgagcaacaaaaaagtg 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 ACTGGAAGTGTGCATCACACAGTTGCTCATCCAGCTGGGAATTTGAGCAACAAAAAAGTG 1680

Qy   1681 acacccactgcgagtactcagcatgatgatgagaatgatactcaaaatggtcttgacaca 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 ACACCCACTGCGAGTACTCAGCATGATGATGAGAATGATACTCAAAATGGTCTTGACACA 1740

Qy   1741 aatatgcataagacagatgtctgtcagcatgtatcagaaatctccacacagaggtgctca 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 AATATGCATAAGACAGATGTCTGTCAGCATGTATCAGAAATCTCCACACAGAGGTGCTCA 1800

Qy   1801 tcaaaggggaaaacagcgggtttgagtaaggggaaaacacattcagctgctagtaccaaa 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 TCAAAGGGGAAAACAGCGGGTTTGAGTAAGGGGAAAACACATTCAGCTGCTAGTACCAAA 1860

Qy   1861 tatggtggtgaaagcaccagaaatggtcagaacatacatgtactcagcgcagaagatcaa 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 TATGGTGGTGAAAGCACCAGAAATGGTCAGAACATACATGTACTCAGCGCAGAAGATCAA 1920

Qy   1921 tgccagatggaaaccgaaaactctgttctgagtcactcggcaaaggtttctccagctgag 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 TGCCAGATGGAAACCGAAAACTCTGTTCTGAGTCACTCGGCAAAGGTTTCTCCAGCTGAG 1980

Qy   1981 catgatatccaaattatgtctgaccttcatgagcagagtctacccaagaagaaaaagaag 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 CATGATATCCAAATTATGTCTGACCTTCATGAGCAGAGTCTACCCAAGAAGAAAAAGAAG 2040

Qy   2041 caaaaacttgaagtgactcgtgaaaaacagaccatgatagatgacatccccatggatatt 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 CAAAAACTTGAAGTGACTCGTGAAAAACAGACCATGATAGATGACATCCCCATGGATATT 2100

Qy   2101 gttgaactgctagctaaaaaccagcatgagaggcagcttatgactgagactgattgttct 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 GTTGAACTGCTAGCTAAAAACCAGCATGAGAGGCAGCTTATGACTGAGACTGATTGTTCT 2160

Qy   2161 gacatcaaccgtattcaatccaagacaactgctgatgatgattgtgtaatagtagctgcc 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 GACATCAACCGTATTCAATCCAAGACAACTGCTGATGATGATTGTGTAATAGTAGCTGCC 2220

Qy   2221 aaggatggttcagattatgcatcaagtgtgtttgacactaattcccaacagaagtccttg 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 AAGGATGGTTCAGATTATGCATCAAGTGTGTTTGACACTAATTCCCAACAGAAGTCCTTG 2280

Qy   2281 gcatcccaaagtacacagaaggagttacagggtcatttggcattgaccacacaagagtct 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 GCATCCCAAAGTACACAGAAGGAGTTACAGGGTCATTTGGCATTGACCACACAAGAGTCT 2340

Qy   2341 ccacatcctcagaactttcagtctactcaggaacagcagacacatttgcggatggaagaa 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 CCACATCCTCAGAACTTTCAGTCTACTCAGGAACAGCAGACACATTTGCGGATGGAAGAA 2400

Qy   2401 atggtcactattgctgcaagctcaccactattttcacatcatgatgatcagtatattgct 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 ATGGTCACTATTGCTGCAAGCTCACCACTATTTTCACATCATGATGATCAGTATATTGCT 2460

Qy   2461 gaagcaccaactgaacattggggccgtaaggacgcaaagaagctaacgtgggagcaattt 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 GAAGCACCAACTGAACATTGGGGCCGTAAGGACGCAAAGAAGCTAACGTGGGAGCAATTT 2520

Qy   2521 aaggccactacaagaaattctccagcagcaacatgtggtgctcaatttagacctggtatc 2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521 AAGGCCACTACAAGAAATTCTCCAGCAGCAACATGTGGTGCTCAATTTAGACCTGGTATC 2580

Qy   2581 caagcagttgacttgacttctactcatgtcatgggatcttccagcaattatgcatctcgc 2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 CAAGCAGTTGACTTGACTTCTACTCATGTCATGGGATCTTCCAGCAATTATGCATCTCGC 2640

Qy   2641 caaccagtaattgcgccactggaccgctatgctgaaagagcggttaaccaggtccatgca 2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2641 CAACCAGTAATTGCGCCACTGGACCGCTATGCTGAAAGAGCGGTTAACCAGGTCCATGCA 2700

Qy   2701 agaaattttccaagcacaatagcaaccatggaagcgagtaagttatgtgatcggagaaat 2760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2701 AGAAATTTTCCAAGCACAATAGCAACCATGGAAGCGAGTAAGTTATGTGATCGGAGAAAT 2760

Qy   2761 gctggacaagtagtcttgtatcctaaagaatccatgcctgcgacgcatcttctgagaatg 2820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2761 GCTGGACAAGTAGTCTTGTATCCTAAAGAATCCATGCCTGCGACGCATCTTCTGAGAATG 2820

Qy   2821 atggatccatcaacattagcaagcttccccaactatggaacttctagcaggaaccagatg 2880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2821 ATGGATCCATCAACATTAGCAAGCTTCCCCAACTATGGAACTTCTAGCAGGAACCAGATG 2880

Qy   2881 gagtctcaacttcataattctcagtatgcacataatcagtacaaaggatcaaccagcaca 2940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2881 GAGTCTCAACTTCATAATTCTCAGTATGCACATAATCAGTACAAAGGATCAACCAGCACA 2940

Qy   2941 tcatatggcagtaacctgaatggaaagattccattgacattcgaagacttatcacggcat 3000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2941 TCATATGGCAGTAACCTGAATGGAAAGATTCCATTGACATTCGAAGACTTATCACGGCAT 3000

Qy   3001 cagctgcatgatctgcacagacctttacgcccacatcctagagttggtgtgcttggctcc 3060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3001 CAGCTGCATGATCTGCACAGACCTTTACGCCCACATCCTAGAGTTGGTGTGCTTGGCTCC 3060

Qy   3061 ttgctgcagaaggaaattgcaaactggtcggagaactgtggcacacaatctggttataag 3120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3061 TTGCTGCAGAAGGAAATTGCAAACTGGTCGGAGAACTGTGGCACACAATCTGGTTATAAG 3120

Qy   3121 ttaggagtgtcaacaggaataacatcgcatcagatgaacagaaaggaacattttgaagcc 3180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3121 TTAGGAGTGTCAACAGGAATAACATCGCATCAGATGAACAGAAAGGAACATTTTGAAGCC 3180

Qy   3181 ctgaattctggaatgttttcagcaaaatggaatgcattgcagttgggttctgttagctcc 3240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3181 CTGAATTCTGGAATGTTTTCAGCAAAATGGAATGCATTGCAGTTGGGTTCTGTTAGCTCC 3240

Qy   3241 agtgcagatttttatcagcgaggaacagcatagctcaatcttggaccagaggcaagggt 3300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3241 AGTGCATATTTTTTATCAGCGAGGAACAGCATAGCTCAATCTTGGACCAGAGGCAAGGGT 3300

Qy   3301 aaaatggttcatcccttggatcggtttgtgagacaggatatctgtataactaacaagaac 3360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3301 AAAATGGTTCATCCCTTGGATCGGTTTGTGAGACAGGATATCTGTATAACTAACAAGAAC 3360

Qy   3361 ccagctgattttactacaatcagtaacgataacgagtatatggattaccgctgaagcaga 3420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3361 CCAGCTGATTTTACTACAATCAGTAACGATAACGAGTATATGGATTACCGCTGAAGCAGA 3420

Qy   3421 aagtggtgtgcataattcctgaacatttacaatcatacatttcatctttatggcgccaaa 3480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3421 AAGTGGTGTGCATAATTCCTGAACATTTACAATCATACATTTCATCTTTATGGCGCCAAA 3480

Qy   3481 tagtcatactgtaagaggagggctttgctggatctgctgtaaggcttcttgtaagttgtg 3540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3481 TAGTCATACTGTAAGAGGAGGGCTTTGCTGGATCTGCTGTAAGGCTTCTTGTAAGTTGTG 3540

Qy   3541 gatgccccatttttctggatgggaacctgccagacagtgaacaagggctttgcaaggtgca 3600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3541 GATGCCCCATTTTCTGGATGGGAACCTGCCAGACAGTGAACAAGGGCTTTGCAAGGTGCA 3600

Qy   3601 gcatccggtttttgttttgccagtccaagaaacgtcctcctgttactttgtagttgtact 3660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3601 GCATCCGGTTTTTGTTTTGCCAGTCCAAGAAACGTCCTCCTGTTACTTTGTAGTTGTACT 3660

Qy   3661 catactagtgcgcttgtttgtacaaggagaaatgtgtaaccttgttgaaaaaaatgtctc 3720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3661 CATACTAGTGCGCTTGTTTGTACAAGGAGAAATGTGTAACCTTGTTGAAAAAAATGTCTC 3720

Qy   3721 ccccattttgtaattaccataaggaggtttatagtgttgtgagctgtgtgtgactgacgg 3780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3721 CCCCATTTTGTAATTACCATAAGGAGGTTTATAGTGTTGTGAGCTGTGTGTGACTGACGG 3780

Qy   3781 cgagaaatggttttgtcggtgttaaggttgaaacgactagctctcgttatcaatgtgttg 3840
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3781 CGAGAAATGGTTTTGTCGGTGTTAAGGTTGAAACGACTAGCTCTCGTTATCAATGTGTTG 3840

Qy    3841 taaacttctagattgatgtgttaccttactcttgaagtcaacaccggagaatttac 3896
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3841 TAAACTTCTAGATTGATGTGTTACCTTACTCTTGAAGTCAACACCGGAGAATTTAC 3896


RESULT   2
AP001859
LOCUS       AP001859               150594 bp    DNA     linear   PLN 27-MAY-2000
DEFINITION  Oryza sativa genomic DNA, chromosome 1, PAC clone:P0485D09.
ACCESSION   AP001859
VERSION     AP001859.1  GI:7630233
KEYWORDS    .
SOURCE      Oryza sativa (cultivar:Nipponbare) DNA, clone:P0485D09.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 150594)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
            clone:P0485D09
  JOURNAL   Published Only in DataBase (2000) In press
REFERENCE   2  (bases 1 to 150594)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-APR-2000) Takuji Sasaki, National Institute of
            Agrobiological Resources, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@abr.affrc.go.jp,
            URL:http://www.dna.affrc.go.jp:82/, Tel:81-298-38-7441,
            Fax:81-298-38-7468)
COMMENT     The orientation of the sequence is from T7 to SP6 of the PAC clone.
            Genes were predicted from the integrated results of the
            following:GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as
            SplicePredictor (October 1998 version). The genomic sequence was
            searched against the non-redundant database NRP (PIR, SWISSPROT,
            GENPEPT, PDB) from MAFF DNA bank and the cDNA sequence database at
            RGP. Protein similarities of the coding regions were searched
            against NRP with BLASTP2.0. ESTs represent the identified cDNA
            sequences using BLASTN2.0 with the corresponding DDBJ accession no.
            and RGP clone ID.
            This sequence of this clone has an overlap with P0431F01 clone,
            DDBJ:AP001550 at the 3' end. This clone ends at the position 57,376
            of P0431F01. Detailed information on overlap and assembly quality
            together with annotation of this entry at
            http://www.dna.affrc.go.jp:82/genomicdata/GenomeFinished.html.
FEATURES              Location/Qualifiers
     source           1. .150594
                      /organism="Oryza sativa"
                      /cultivar="Nipponbare"
                      /db_xref="taxon:4530"
                      /chromosome="1"
                      /clone="P0485D09"
     CDS              join(7919. .7964,8092. .8280,8453. .8473,8558. .8610)
                      /note="hypothetical protein"
                      /codon_start=1
                      /protein_id="BAA94767.1"
                      /db_xref="GI:7630234"
                      /translation="METRIFIQVRAPELSVDNRIAFLLESVPGEIRDSAFVSPDSIRR
                      HRRVLAVLIPEVDIRRRVLAYVGFYVVASMSIVLGCFRIGHVRSPNPGQIRLQQIKAK
                      "
     CDS              complement(9235. .9582)
                      /note="hypothetical protein"
                      /codon_start=1
                      /protein_id="BAA94768.1"
                      /db_xref="GI:7630235"
                      /translation="MTARFAEVGERREAPAGRSMEEGGGGAHRRRGASPAAAATRIGG
                      ARRNFGREGKMRRGVDGYLAIFPGWFLQKGRGREGARGCEMDGGEMARVGSARSGPSV
                      STSAAFVSRFGNF"
     CDS              join(10735. .10850,10933. .10987,11596. .11653,12285. .12379,
                      12471. .12574,12958. .13086,13493. .15302)
                      /note="Similar to Arabidopsis thaliana chromosome 4, BAC
                      clone F4D11; putative myb-protein. (AL022537)"
                      /codon_start=1
                      /protein_id="BAA94769.1"
                      /db_xref="GI:7630236"
                      /translation="MTSDNGKAPDKDGEPSGPPSAPQEGEISNEPKRRRPLNGRTTGP
                      TRRSTKGNWTPEEDAILSRAVQTYNGKNWKKIAECFPDRTDVQCLHRWQKVLNPELVK
                      GPWSKEEDEIIVQMVNKLGPKKWSTIAQALPGRIGKQCRERWYNHLNPGINKEAWTQE
                      EEITLIHAHRMYGNKWAELTKFLPGRTDNSIKNHWNSSVKKKVNSYMSSGLLTQVSCL
                      PLNEYSANCNSSPAMTQQNSEDSGCFAVREVENSSGCSQSSLAKVSCSQVHDTTVPLG
                      CDLQVNANFDKNEAHDSQSSMGPQACYTSAEAVASALPAVHCHVSSSNLDPDQHLQED
                      FAQGLNLDMTIDEMPTVPSFADNQTVCSIENHERSLEPYDVAMEVPLSMLSSDSGAEQ
                      KLHFMSEADFNSPNCLKSELWQDISLQGLLSGPDAVEADSISRSNHQSDVYSSEADTH
                      FLAPPYMPQTSNSSSVMGLADDQSPQMSVPPSLICSNAMTDDAPFDNRPGRKEMPLSQ
                      AEVVTQSSSSSGDAEMFANPGCSNDRHVPSSTMESIPECGDQQVTNAEEPEASLEKEP
                      SLTQSVTAPDEQDKGALFYEPPRFPSLDVPFVSCDLVTSGDLQEFSPLGIRQLMHSTM
                      NVCTPMRLWGSPTHDESTGVLLKSAAKSFICTPSILKKRHRDLLSPIPDKRIEKKYGT
                      EKDRGVSDTSSTGIQTSCINATKDDALITTVLRIERSASSKSLEKKLVFSDENKENLG
                      YTTEQTKDGQSAGNDEHMDEQTTGERSSATNVATNDDLSGNLVSTSSF"
     CDS              complement(join(18608. .18823,19905. .20031,20072. .20220,
                      20774. .21113,21407. .21738,21975. .22508))
                      /note="ESTs AU082761(S5084),D42006(S5084) correspond to a
                      region of the predicted gene.
                      Similar to Caenorhabditis elegans chromosome 3, cosmid
                      C17G10; alcohol dehydrogenase/ribitol dehydrogenase.
                      (U28739)"
                      /codon_start=1
                      /protein_id="BAA94770.1"
                      /db_xref="GI:7630237"
                      /translation="MAAAAAEGGGGGGGAVYDPSYVPDSVKTFVAHMYRHVRDKNVYE
                      IHQMYEGGFQRLSDRLFRDAPWPAAEAVSPYCDGDHVFLLLYRELWYRHAYARASSSS
                      SSSAPLTAGQRAESWANYCDLFSVVLHGVVNMQLPNQWLWDMVDEFVYQFQSFCQYRA
                      KLKNKSDDELHQLKQFDKAWNVYGVLNYLQALVEKSMIAQILEREKEGLEQFTATDGY
                      DYQGGSNVLKMLGYYSMIGLLRIHCLLGDYRTGLKCLAPIDLNQQGVYTIVIGSHISA
                      IYHYGFANLMMRRYAEAIREFNKILLYILKYKQYHQKSPQYDQILKKNEQMYAFLAVC
                      LSLCPQHNLIDENVSTQLKEKYNDKMTKMQRFDEETYAAYDELFSYACPKFITPSPPA
                      LDQPLTNYNQDAYRLQLKLFLYEVKQQQLLSGIRSYLKLYSTITIAKLAQYMEVDEAT
                      LRAMKNLSQYLYRSILMTYKHKMHAVDNNGKIVSSADFDFYIKEICSSIRTPKLTSKN
                      LAHIHIQVGVKTVRKPSLSFSFSYFFFGIEIGNSGYENGIEYYRNRKWSENEPAQIR"
     CDS              join(23045. .23145,24592. .24670,25118. .25216,25302. .25415,
                      25608. .25750,26364. .26445,27073. .27158,27527. .27760,
                      29332. .29917)
                      /note="Similar to Arabidopsis thaliana chromosome 2 BAC
                      T19C21; unknown protein. (AC004683)"
                      /codon_start=1
                      /protein_id="BAA94771.1"
                      /db_xref="GI:7630238"
                      /translation="MAGSMVDRATSDMLIGPDWAKNMEICDICNRDPGQSKDVVKALK
                      KRIGHKNPKVQILALTLLETAIKNCGDIFHMHVAERDVLHEMVKIVKKKSDQNVKEKV
                      LTMIDTWQEAFGGPRARYPQYYAAYHDLVRAGAAFPKRSDRPAPLFNGQSPAGRNMRS
                      PDQQDEAESSAGNDFPALSMSEIQNARGIMDVLAEMLNALDPGNREGLRQEVIVELVD
                      QCRTYKQRVVLLVNATADEELMSQGLALNDDLQRVLAKHDAIAAGIAVRVEKKPKSLQ
                      ALVDTEDSMNQDSKKEQALVDIEDPTTQETNKEPSQSASVQSPFEQLALPAPPVSNGS
                      ATPAPKSDLGVDLLSWDDNPSTTENSLALVPVTDPVVDSTPSQNALAIVDIFSQNNTT
                      NNIAKPADPFGVNSSSTLPGSQPYNAPTQHPLQAQQQPQQVGLYPNGGAVNPGTSYPT
                      SSGWNGQIANNATPPAQQTVNYGIAIYLYIYISQPFSYTIIPIIGCQNHACPCLCSV"
     CDS              30166. .30675
                      /note="ESTs AU031545(E61825),AU031546(E61825) correspond
                      to a region of the predicted gene.
                      hypothetical protein"
                      /codon_start=1
                      /protein_id="BAA94772.1"
                      /db_xref="GI:7630239"
                      /translation="MYNNQQPGAMQPSQPAVTQMQSGFANQFGSMPPHSMPGMQFPGM
                      QPSPMPGAQPVMMYAQPMMMPGMQFAAMPQPRMYGPQMSQYRLVQQQAAQYYSNSQGR
                      PTYYAGMNDLSQKMYGLSMQDSSYMGMNSSPYSTTPSSSSSMGQPSKPEDKLFGDLLS
                      IAKTKQNKA"
     CDS              complement(join(34629. .34795,36831. .36867,38803. .38895,
                      41780. .41830))
                      /note="hypothetical protein"
                      /codon_start=1
                      /protein_id="BAA94773.1"
```

```
                /db_xref="GI:7630240"
                /translation="MTVVRCTTRQRLLEGERIQCDIRGKKILGSKGPFDSQNRKNTGI
                EIEVGLLDACVIVYTVEKPENERRIKKHGRLRGLSRLCDRLGPKLVWDPSPTKVVTPY
                VQFPRGHTLFHRG"
     CDS        join(46367. .46423,47537. .47769,48031. .48153,50082. .50299,
                51762. .51864,52091. .52204,52433. .52532,52784. .53004,
                54084. .54514,54601. .55709,56020. .57468)
                /note="hypothetical protein"
                /codon_start=1
                /protein_id="BAA94774.1"
                /db_xref="GI:7630241"
                /translation="MPNGLIRKLREELLDKLIAHATFPHPPPQIPVGAAAAAATDAIA
                WICAPVVAAAWICAPASPVPDGVSPSCLRAVPPPPNRRQTRPSVVAPSPTIDRHIPLL
                IGCFMEKDLIFMTHGRGVVKADEEVRGSRSARTELKIRGVGLMLVDCGYPPIHVGVDG
                FMCVRAVLHRMGSNGGRLALLTCSLTKANCLLAMKLFLFKEKQTFELGNIEEAAASSG
                ERSTARRVDRAPPRLRRGCRSKRKPPSCAPLSGLELLASMGTKALIFHVAEEKKIPSF
                AGILAATFWGNLFLAIILPARISGALIFHELIEVDIYCACLCVDSLLVVVIDMEIVAV
                DQEGARVVGTNCMLARGGTGAVAPVLELTATPRQDAAAEAGVDEPAQHQCEHFSIRGY
                VALLQKKDPKFCSLSRIFHDQKKCDEHKASSSPFSVAKFRRWDCSKCLDKLKTSDNGT
                APRTLPAKQNGTSDGCSITFVRSTFVPASVGSQKVSPSTQSSQGKNADRSTLPKSVQE
                GNDSKCNAPSGKNGAAEANTDSPMKDLQGPAQNYDVAANVSEDNTSVDVGALPEVPQI
                TWHIEVNGADQPPSTPKLSEVVLKRNEDENGKTEETLVAEQCNLTKDPNPMSGKERDQ
                VAEQCNLTKDPKPVSGQKCEQICNEPCEEVVLKRSSKSKRKTDKKLMKKQQHSKKRTA
                QADVSDAKLCRRKPKKVRLLSEIINANQVEDSRSDEVHRENAADPCEDDRSTIPVPME
                VSMDIPVSNHTVGEDGLKSSKNKTKRKYSDVVDDGSSLMNWLNGKKKRTGSVHHTVAH
                PAGNLSNKKVTPTASTQHDDENDTENGLDTNMHKTDVCQHVSEISTQRCSSKGKTAGL
                SKGKTHSAASTKYGGESTRNGQNIHVLSAEDQCQMETENSVLSHSAKVSPAEHDIQIM
                SDLHEQSLPKKKKKQKLEVTREKQTMIDDIPMDIVELLAKNQHERQLMTETDCSDINR
                IQSKTTADDDCVIVAAKDGSDYASSVFDTNSQQKSLASQSTQKELQGHLALTTQESPH
                PQNFQSTQEQQTHLRMEEMVTIAASSPLFSHHDDQYIAEAPTEHWGRKDAKKLTWEQF
                KATTRNSPAATCGAQFRPGIQAVDLTSTHVMGSSSNYASRQPVIAPLDRYAERAVNQV
                HARNFPSTIATMEASKLCDRRNAGQVVLYPKESMPATHLLRMMDPSTLASFPNYGTSS
                RNQMESQLHNSQYAHNQYKGSTSTSYGSNLNGKIPLTFEDLSRHQLHDLHRPLRPHPR
                VGVLGSLLQKEIANWSENCGTQSGYKLGVSTGITSHQMNRKEHFEALNSGMFSAKWNA
                LQLGSVSSSADFLSARNSIAQSWTRGKGKMVHPLDRFVRQDICITNKNPADFTTISND
                NEYMDYR"
     CDS        complement(join(58634. .58667,58796. .58902,59018. .59083,
                59368. .59432,59585. .59648,60333. .60442,60620. .60707,
                60835. .60945))
                /note="ESTs AU082698(R0665),D23963(R0665) correspond to a
                region of the predicted gene.
                Similar to Arabidopsis thaliana rac GTP binding protein
                Arac7. (AF079484)"
                /codon_start=1
                /protein_id="BAA94775.1"
                /db_xref="GI:7630242"
                /translation="MSSAAAATRFIKCVTVGDGAVGKTCMLICYTCNKFPTDYIPTVF
                DNFSANVSVDGSVVNLGLWDTAGQEDYSRLRPLSYRGADVFILSFSLISRASYENVQK
                KWMPELRRFAPGVPVVLVGTKLDLREDRAYLADHPASSIITTEQGEELRKLIGAVAYI
                ECSSKTQRNIKAVFDTAIKVVLQPPRHKDVTRKKLQSSSNRPVRRYFCGSACFA"
     CDS        join(67330. .67530,67676. .67816,69073. .69234)
                /note="hypothetical protein"
                /codon_start=1
                /protein_id="BAA94776.1"
                /db_xref="GI:7630243"
                /translation="MGGSAAAAPPAPAWMAAGARRWLEEAGVAFDGSDRRAFNALPLA
                GARVSLAEAGRAVCSLRVTAELTDAEGNWHPGAIAAAADDVCAAAIMSVEGIIKVSVH
                YDISYFSPAKLHEEVELDGRVVEQKGKMTAVTVEIRKKDSGELVAIGRQWMSTTRPKK
                DQASSKL"
     CDS        join(69755. .69967,70096. .70236,70344. .70514)
                /note="Similar to Arabidopsis thaliana chromosome 1 BAC
                F19P19; unknown protein (AC000104)"
                /codon_start=1
                /protein_id="BAA94777.1"

  Query Match             75.8%;  Score 2952.2;  DB 8;  Length 150594;
  Best Local Similarity   87.4%;  Pred. No. 0;
  Matches 3487;  Conservative    0;  Mismatches   23;  Indels  481;  Gaps    4;

Qy     386 acgaaccggcacaacaccaatgcgagcatttctccataagagggtatgttgctcttcttc 445
           | | | |   | ||    |||   |   ||  |  | ||||||||||||||||||||||
Db   54044 ATGCATCTTTATAATGATAATCTTATTTTTCTTGTACCAGAGGGTATGTTGCTCTTCTTC 54103

Qy     446 agaagaaggatccaaaattctgctctctatctcggattttccatgaccagaaaaaatgtg 505
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   54104 AGAAGAAGGATCCAAAATTCTGCTCTCTATCTCGGATTTTCCATGACCAGAAAAAATGTG 54163

Qy     506 atgaacacaaagctagttcaagcccatttttctgtagcaaagtttcgacgatgggattgct 565
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   54164 ATGAACACAAAGCTAGTTCAAGCCCATTTTCTGTAGCAAAGTTTCGACGATGGGATTGCT 54223

Qy     566 cgaagtgcttggataagttgaaaacttcagataatggaacagcaccaagaactcttcccg 625
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   54224 CGAAGTGCTTGGATAAGTTGAAAACTTCAGATAATGGAACAGCACCAAGAACTCTTCCCG 54283

Qy     626 caaagcagaatggcacaagtgatggttgctccatcacatttgttcggagcacttttgtgc 685
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   54284 CAAAGCAGAATGGCACAAGTGATGGTTGCTCCATCACATTTGTTCGGAGCACTTTTGTGC 54343

Qy     686 ctgctagtgttggttcccaaaaagtgtctcctagcacacaatcatctcaagggaagaatg 745
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   54344 CTGCTAGTGTTGGTTCCCAAAAAGTGTCTCCTAGCACACAATCATCTCAAGGGAAGAATG 54403

Qy     746 ctgatagatcaactcttccaaagagtgtgcaagaaggcaatgactccaaatgcaatgcgc 805
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   54404 CTGATAGATCAACTCTTCCAAAGAGTGTGCAAGAAGGCAATGACTCCAAATGCAATGCGC 54463

Qy     806 cttctggcaagaatggagctgctgaggccaatactgattcaccaatgaa---------- 854
           ||||||||||||||||||||||||||||||||||||||||||||||||||          
Db   54464 CTTCTGGCAAGAATGGAGCTGCTGAGGCCAATACTGATTCACCAATGAAAGGTATGGTAG 54523

Qy     855 ---------------------------------------------------------- 854

Db   54524 ATGTAGAGCCTTTCAAATTCCTAAGTAGGATTTTATTTAAGGTATAGAATAAACTAATGT 54583

Qy     855 ----------agatttgcaagggccagcccaaaattatgatgtggcagcaaatgt 899
                      ||||||||||||||||||||||||||||||||||||||||||||
Db   54584 TTGTGTGATTTTCTCAGATTTGCAAGGGCCAGCCCAAAATTATGATGTGGCAGCAAATGT 54643

Qy     900 ctctgaggacaacacttctgttgatgttggggctttacctgaagttccccagattacatg 959
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   54644 CTCTGAGGACAACACTTCTGTTGATGTTGGGGCTTTACCTGAAGTTCCCCAGATTACATG 54703

Qy     960 gcacatagaagtaaatggtgcagatcaacctccatccactccaaaactttctgaagtggt 1019
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   54704 GCACATAGAAGTAAATGGTGCAGATCAACCTCCATCCACTCCAAAACTTTCTGAAGTGGT 54763

Qy    1020 cctcaaaagaaatgaagatgaaaatggaaaaactgaagagactcttgttgctgagcagtg 1079
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   54764 CCTCAAAAGAAATGAAGATGAAAATGGAAAAACTGAAGAGACTCTTGTTGCTGAGCAGTG 54823

Qy    1080 caatttgaccaaagatcctaacccaatgtctggaaaggaacgtgatcaggttgctgagca 1139
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   54824 CAATTTGACCAAAGATCCTAACCCAATGTCTGGAAAGGAACGTGATCAGGTTGCTGAGCA 54883

Qy    1140 gtgcaatttgaccaaagatccgaaaccagtgtctgggcagaaatgtgagcagatctgcaa 1199
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   54884 GTGCAATTTGACCAAAGATCCGAAACCAGTGTCTGGGCAGAAATGTGAGCAGATCTGCAA 54943

Qy    1200 tgagccatgtgaagaagttgttctcaaaagaagctccaaatctaagaggaagacggataa 1259
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   54944 TGAGCCATGTGAAGAAGTTGTTCTCAAAAGAAGCTCCAAATCTAAGAGGAAGACGGATAA 55003

Qy    1260 gaagttgatgaagaagcagcagcacagcaagaaacgcactgcccaggctgatgtttcaga 1319
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   55004 GAAGTTGATGAAGAAGCAGCAGCACAGCAAGAAACGCACTGCCCAGGCTGATGTTTCAGA 55063

Qy    1320 tgcaaagctttgtcggagaaagccaaaaaaggtgcggcttctatcagaaattataaatgc 1379
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   55064 TGCAAAGCTTTGTCGGAGAAAGCCAAAAAAGGTGCGGCTTCTATCAGAAATTATAAATGC 55123

Qy    1380 taaccaggttgaggattctagaagtgacgaagttcatcgtgaaaatgccgctgatccctg 1439
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   55124 TAACCAGGTTGAGGATTCTAGAAGTGACGAAGTTCATCGTGAAAATGCCGCTGATCCCTG 55183

Qy    1440 tgaggatgatagaagtaccatcccggtcccgatggaagtaagcatggatattcctgttag 1499
```

```
Db  55184 TGAGGATGATAGAAGTACCATCCCGGTCCCGATGGAAGTAAGCATGGATATTCCTGTTAG 55243

Qy   1500 caaccatacagtgggagaagatgggttaaaatcaagtaagaacaagacaaaacgcaaata 1559

Db  55244 CAACCATACAGTGGGAGAAGATGGGTTAAAATCAAGTAAGAACAAGACAAAACGCAAATA 55303

Qy   1560 ctctgatgttgtagatgatggatcatcacttatgaactggctgaatggaaaaaagaaaag 1619

Db  55304 CTCTGATGTTGTAGATGATGGATCATCACTTATGAACTGGCTGAATGGAAAAAAGAAAAG 55363

Qy   1620 aactggaagtgtgcatcacacagttgctcatccagctgggaatttgagcaacaaaaaagt 1679

Db  55364 AACTGGAAGTGTGCATCACACAGTTGCTCATCCAGCTGGGAATTTGAGCAACAAAAAAGT 55423

Qy   1680 gacacccactgcgagtactcagcatgatgatgagaatgatactgaaaatggtcttgacac 1739

Db  55424 GACACCCACTGCGAGTACTCAGCATGATGATGAGAATGATACTGAAAATGGTCTTGACAC 55483

Qy   1740 aaatatgcataagacagatgtctgtcagcatgtatcagaaatctccacacagaggtgctc 1799

Db  55484 AAATATGCATAAGACAGATGTCTGTCAGCATGTATCAGAAATCTCCACACAGAGGTGCTC 55543

Qy   1800 atcaaagggaaaacagcgggtttgagtaaggggaaaacacattcagctgctagtaccaa 1859

Db  55544 ATCAAAGGGGAAAACAGCGGGTTTGAGTAAGGGGAAAACACATTCAGCTGCTAGTACCAA 55603

Qy   1860 atatggtggtgaaagcaccagaaatggtcagaacatacatgtactcagcgcagaagatca 1919

Db  55604 ATATGGTGGTGAAAGCACCAGAAATGGTCAGAACATACATGTACTCAGCGCAGAAGATCA 55663

Qy   1920 atgccagatggaaaccgaaaactctgttctgagtcactcggcaa---------------- 1963

Db  55664 ATGCCAGATGGAAACCGAAAACTCTGTTCTGAGTCACTCGGCAAAGGTACGAATTTTGTG 55723

Qy   1964 ------------------------------------------------------------ 1963

Db  55724 AATCATGAGGAATTTTTGCTTTTTAAATTGACTGAATCAACATTTATCTGTATGAAGGAA 55783

Qy   1964 ------------------------------------------------------------ 1963

Db  55784 TAATATTGGTGCATAACAATGTTAAGAAATATGCATACAATGTTTATTTATATGCTTTCC 55843

Qy   1964 ------------------------------------------------------------ 1963

Db  55844 ACTGTTCTTCTTTACTTATGTTTTGATACTCTTTTTGTGTGTGCGTGCATGTGTGCATGT 55903

Qy   1964 ------------------------------------------------------------ 1963

Db  55904 GTGTGTGTGTGTGTGTGTGCGCGCGTGTGTGTGCACGTGCGTGGCGCAATATTCTTTTTT 55963

Qy   1964 ------------------------------------------------------------aggttt 1969

Db  55964 AGACTCATATTATAGTGATTGTAATGGACTGACATTTTCCTCATTTCTCATCTCAGGTTT 56023

Qy   1970 ctccagctgagcatgatatccaaattatgtctgaccttcatgagcagagtctacccaaga 2029

Db  56024 CTCCAGCTGAGCATGATATCCAAATTATGTCTGACCTTCATGAGCAGAGTCTACCCAAGA 56083

Qy   2030 agaaaaagaagcaaaaacttgaagtgactcgtgaaaaacagaccatgatagatgacatcc 2089

Db  56084 AGAAAAAGAAGCAAAAACTTGAAGTGACTCGTGAAAAACAGACCATGATAGATGACATCC 56143

Qy   2090 ccatggatattgttgaactgctagctaaaaaccagcatgagaggcagcttatgactgaga 2149

Db  56144 CCATGGATATTGTTGAACTGCTAGCTAAAAACCAGCATGAGAGGCAGCTTATGACTGAGA 56203

Qy   2150 ctgattgttctgacatcaaccgtattcaatccaagacaactgctgatgatgattgtgtaa 2209

Db  56204 CTGATTGTTCTGACATCAACCGTATTCAATCCAAGACAACTGCTGATGATGATTGTGTAA 56263

Qy   2210 tagtagctgccaaggatggttcagattatgcatcaagtgtgtttgacactaattcccaac 2269

Db  56264 TAGTAGCTGCCAAGGATGGTTCAGATTATGCATCAAGTGTGTTTGACACTAATTCCCAAC 56323

Qy   2270 agaagtccttggcatcccaaagtacacagaaggagttacagggtcatttggcattgacca 2329

Db  56324 AGAAGTCCTTGGCATCCCAAAGTACACAGAAGGAGTTACAGGGTCATTTGGCATTGACCA 56383

Qy   2330 cacaagagtctccacatcctcagaactttcagtctactcaggaacagcagacacatttgc 2389

Db  56384 CACAAGAGTCTCCACATCCTCAGAACTTTCAGTCTACTCAGGAACAGCAGACACATTTGC 56443

Qy   2390 ggatggaagaaatggtcactattgctgcaagctcaccactattttcacatcatgatgatc 2449

Db  56444 GGATGGAAGAAATGGTCACTATTGCTGCAAGCTCACCACTATTTTCACATCATGATGATC 56503

Qy   2450 agtatattgctgaagcaccaactgaacattggggccgtaaggacgcaaagaagctaacgt 2509

Db  56504 AGTATATTGCTGAAGCACCAACTGAACATTGGGGCCGTAAGGACGCAAAGAAGCTAACGT 56563

Qy   2510 gggagcaatttaaggccactacaagaaattctccagcagcaacatgtggtgctcaattta 2569

Db  56564 GGGAGCAATTTAAGGCCACTACAAGAAATTCTCCAGCAGCAACATGTGGTGCTCAATTTA 56623

Qy   2570 gacctggtatccaagcagttgacttgacttctactcatgtcatgggatcttccagcaatt 2629

Db  56624 GACCTGGTATCCAAGCAGTTGACTTGACTTCTACTCATGTCATGGGATCTTCCAGCAATT 56683

Qy   2630 atgcatctcgccaaccagtaattgcgccactggaccgctatgctgaaagagcggttaacc 2689

Db  56684 ATGCATCTCGCCAACCAGTAATTGCGCCACTGGACCGCTATGCTGAAAGAGCGGTTAACC 56743

Qy   2690 aggtccatgcaagaaattttccaagcacaatagcaaccatggaagcgagtaagttatgtg 2749

Db  56744 AGGTCCATGCAAGAAATTTTCCAAGCACAATAGCAACCATGGAAGCGAGTAAGTTATGTG 56803

Qy   2750 atcggagaaatgctggacaagtagtcttgtatcctaaagaatccatgcctgcgacgcatc 2809

Db  56804 ATCGGAGAAATGCTGGACAAGTAGTCTTGTATCCTAAAGAATCCATGCCTGCGACGCATC 56863

Qy   2810 ttctgagaatgatggatccatcaacattagcaagcttccccaactatggaacttctagca 2869

Db  56864 TTCTGAGAATGATGGATCCATCAACATTAGCAAGCTTCCCCAACTATGGAACTTCTAGCA 56923

Qy   2870 ggaaccagatggagtctcaacttcataattctcagtatgcacataatcagtacaaaggat 2929

Db  56924 GGAACCAGATGGAGTCTCAACTTCATAATTCTCAGTATGCACATAATCAGTACAAAGGAT 56983

Qy   2930 caaccagcacatcatatggcagtaacctgaatggaaagattccattgacattcgaagact 2989

Db  56984 CAACCAGCACATCATATGGCAGTAACCTGAATGGAAAGATTCCATTGACATTCGAAGACT 57043

Qy   2990 tatcacggcatcagctgcatgatctgcacagacctttacgcccacatcctagagttggtg 3049

Db  57044 TATCACGGCATCAGCTGCATGATCTGCACAGACCTTTACGCCCACATCCTAGAGTTGGTG 57103

Qy   3050 tgcttggctccttgctgcagaaggaaattgcaaactggtcggagaactgtggcacacaat 3109

Db  57104 TGCTTGGCTCCTTGCTGCAGAAGGAAATTGCAAACTGGTCGGAGAACTGTGGCACACAAT 57163

Qy   3110 ctggttataagttaggagtgtcaacaggaataacatcgcatcagatgaacagaaaggaac 3169

Db  57164 CTGGTTATAAGTTAGGAGTGTCAACAGGAATAACATCGCATCAGATGAACAGAAAGGAAC 57223

Qy   3170 attttgaagccctgaattctggaatgttttcagcaaaatggaatgcattgcagttgggtt 3229

Db  57224 ATTTTGAAGCCCTGAATTCTGGAATGTTTTCAGCAAAATGGAATGCATTGCAGTTGGGTT 57283

Qy   3230 ctgttagctccagtgcagatttttttatcagcgaggaacagcatagctcaatcttggacca 3289

Db  57284 CTGTTAGCTCCAGTGCAGATTTTTTTATCAGCGAGGAACAGCATAGCTCAATCTTGGACCA 57343

Qy   3290 gaggcaagggtaaaatggttcatcccttggatcggtttgtgagacaggatatctgtataa 3349

Db  57344 GAGGCAAGGGTAAAATGGTTCATCCCTTGGATCGGTTTGTGAGACAGGATATCTGTATAA 57403
```

```
Qy   3350 ctaacaagaacccagctgattttactacaatcagtaacgataacgagtatatggattacc 3409
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  57404 CTAACAAGAACCCAGCTGATTTTACTACAATCAGTAACGATAACGAGTATATGGATTACC 57463

Qy   3410 gctgaagcagaaagtggtgtgcataattcctgaacatttacaatcatacatttcatcttt 3469
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  57464 GCTGAAGCAGAAAGTGGTGTGCATAATTCCTGAACATTTACAATCATACATTTCATCTTT 57523

Qy   3470 atggcgccaaatagtcatactgtaagaggagggctttgctggatctgctgta-------- 3521
          ||||||||||||||||||||||||||||||||||||||||||||||||||||        |
Db  57524 ATGGCGCCAAATAGTCATACTGTAAGAGGAGGGCTTTGCTGGATCTGCTGTAAGGTAAGT 57583

Qy   3522 ---------------------------------------------------------- 3521

Db  57584 TGAACTTTTTCTTCTTGCAAGTTTATCAGTTTAAGAAAAAAGAATGATTACTTATGTTAG 57643

Qy   3522 ----------------aggcttcttgtaagttgtggatgccccatttttctggatgggaac 3565
                           ||||||||||||||||||||||||||||||||||||||||||||
Db  57644 CAAGGATGGTTCTTGCAGGCTTCTTGTAAGTTGTGGATGCCCCATTTTCTGGATGGGAAC 57703

Qy   3566 ctgccagacagtgaacaagggctttgcaaggtgcagcatccggtttttgttttgccagtc 3625
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  57704 CTGCCAGACAGTGAACAAGGGCTTTGCAAGGTGCAGCATCCGGTTTTTGTTTTGCCAGTC 57763

Qy   3626 caagaaacgtcctcctgttactttgtagttgtactcatactagtgcgcttgtttgtacaa 3685
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  57764 CAAGAAACGTCCTCCTGTTACTTTGTAGTTGTACTCATACTAGTGCGCTTGTTTGTACAA 57823

Qy   3686 ggagaaatgtgtaaccttgttgaaaaaaatgtctccccccattttgtaattaccataagga 3745
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db  57824 GGAGAAATGTGTAACCTTGTTG-AAAAAATGTCTCCCCCATTTTGTAATTACCATAAGGA 57882

Qy   3746 ggtttatagtgttgtgagctgtgtgtgactgacggcgagaaatggttttgtcggtgttaa 3805
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  57883 GGTTTATAGTGTTGTGAGCTGTGTGTGACTGACGGCGAGAAATGGTTTTGTCGGTGTTAA 57942

Qy   3806 ggttgaaacgactagctctctgttatcaatgtgttgtaaacttctagattgatgtgttacc 3865
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  57943 GGTTGAAACGACTAGCTCTCGTTATCAATGTGTTGTAAACTTCTAGATTGATGTGTTACC 58002

Qy   3866 ttactcttgaagtcaacaccggagaatttac 3896
          ||||||||||||||||||||||||||||||
Db  58003 TTACTCTTGAAGTCAACACCGGAGAATTTAC 58033
```

```
RESULT   3
I66494/c
LOCUS       I66494                  7218 bp    DNA     linear   PAT 28-DEC-1997
DEFINITION  Sequence 14 from patent US 5670367.
ACCESSION   I66494
VERSION     I66494.1  GI:2724471
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 7218)
  AUTHORS   Dorner,F., Scheiflinger,F. and Falkner,F.Gunter.
  TITLE     Recombinant fowlpox virus
  JOURNAL   Patent: US 5670367-A 14 23-SEP-1997;
FEATURES             Location/Qualifiers
     source          1. .7218
                     /organism="unknown"
BASE COUNT     1944 a   1491 c   1486 g   1929 t    368 others
ORIGIN


  Query Match             1.6%;  Score 62.8;  DB 6;  Length 7218;
  Best Local Similarity   7.2%;  Pred. No. 1.2e-05;
  Matches   31;  Conservative  227;  Mismatches  174;  Indels    0;  Gaps    0;

Qy   1182 atgtgagcagatctgcaatgagccatgtgaagaagttgttctcaaaagaagctccaaatc 1241
          ||||  |||   ||| ||| | | || |    |||| |||    | | |
```

```
Db   1497 ATGTAGGCATCACTGTAATTACCTATCTATGCAAGTAGTTAAAGAGATAGAAGAATTTGG 1438

Qy   1242 taagaggaagacggataagaagttgatgaagaagcagcagcacagcaagaaacgcactgc 1301
          || :::::::: ::: :::::  :: :::::::: :: :: : :: :::::: : :  :
Db   1437 TACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1378

Qy   1302 ccaggctgatgtttcagatgcaaagctttgtcggagaaagccaaaaaaggtgcggcttct 1361
            :::  :: :    ::: : ::::    :  :::::::  :::::::: : ::
Db   1377 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1318

Qy   1362 atcagaaattataaatgctaaccaggttgaggattctagaagtgacgaagttcatcgtga 1421
          :  ::::: : ::: :  ::  :::  ::::::    ::::: :: ::::   :  : ::
Db   1317 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1258

Qy   1422 aaatgccgctgatccctgtgaggatgatagaagtaccatcccggtcccgatggaagtaag 1481
          ::: :  :  ::     : ::::: :: ::::: :  :    ::    :: ::::: :::
Db   1257 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1198

Qy   1482 catggatattcctgttagcaaccatacagtgggagaagatgggttaaaatcaagtaagaa 1541
           : ::: :     :  :: ::  : : :: ::::::::: :::  ::::  ::: :::::
Db   1197 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1138

Qy   1542 caagacaaaacgcaaatactctgatgttgtagatgatggatcatcacttatgaactggct 1601
          :::: :::: : ::: :    :: :  : ::: :: :::  :  :   : :::  ::
Db   1137 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1078

Qy   1602 gaatggaaaaaa 1613
          ::: ::::::::|
Db   1077 RRRRRRRRRRRA 1066
```

```
RESULT   4
AC026135
LOCUS       AC026135               141041 bp    DNA     linear   HTG 05-APR-2000
DEFINITION  Homo sapiens chromosome 3 clone RP11-78O22 map 3, WORKING DRAFT
            SEQUENCE, 12 unordered pieces.
ACCESSION   AC026135
VERSION     AC026135.2  GI:7417828
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 141041)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 3, clone RP11-78O22
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 141041)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Apr 5, 2000 this sequence version replaced gi:7264205.
```

```
        All repeats were identified using RepeatMasker:
        Smit, A.F.A. & Green, P. (1996-1997)
        http://ftp.genome.washington.edu/RM/RepeatMasker.html
        ------------- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
        ------------- Project Information
            Center project name: L7353
            Center clone name: 78_O_22
        ------------- Summary Statistics
            Sequencing vector: M13; M77815; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 132847 bases at least Q40
            Consensus quality: 137397 bases at least Q30
            Consensus quality: 139005 bases at least Q20
            Insert size: 145000; agarose-fp
            Insert size: 139941; sum-of-contigs
            Quality coverage: 4.8 in Q20 bases; agarose-fp
            Quality coverage: 4.9 in Q20 bases; sum-of-contigs
        --------------.
        * NOTE: This is a 'working draft' sequence. It currently
        * consists of 12 contigs. The true order of the pieces
        * is not known and their order in this sequence record is
        * arbitrary. Gaps between the contigs are represented as
        * runs of N, but the exact sizes of the gaps are unknown.
        * This record will be updated with the finished sequence
        * as soon as it is available and the accession number will
        * be preserved.
        *        1     2190: contig of 2190 bp in length
        *     2191 2290: gap of      100 bp
        *     2291     5642: contig of 3352 bp in length
        *     5643 5742: gap of      100 bp
        *     5743    10058: contig of 4316 bp in length
        *    10059 10158: gap of      100 bp
        *    10159    14819: contig of 4661 bp in length
        *    14820 14919: gap of      100 bp
        *    14920    19408: contig of 4489 bp in length
        *    19409 19508: gap of      100 bp
        *    19509    26115: contig of 6607 bp in length
        *    26116 26215: gap of      100 bp
        *    26216    39655: contig of 13440 bp in length
        *    39656 39755: gap of      100 bp
        *    39756    54129: contig of 14374 bp in length
        *    54130 54229: gap of      100 bp
        *    54230    64518: contig of 10289 bp in length
        *    64519 64618: gap of      100 bp
        *    64619    79904: contig of 15286 bp in length
        *    79905 80004: gap of      100 bp
        *    80005   109845: contig of 29841 bp in length
        *   109846 109945: gap of      100 bp
        *   109946   141041: contig of 31096 bp in length.
FEATURES             Location/Qualifiers
     source          1. .141041
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /map="3"
                     /clone="RP11-78O22"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .2190
                     /note="assembly_fragment"
     misc_feature    2291. .5642
                     /note="assembly_fragment"
     misc_feature    5743. .10058
                     /note="assembly_fragment"
     misc_feature    10159. .14819
                     /note="assembly_fragment"
     misc_feature    14920. .19408
                     /note="assembly_fragment"
     misc_feature    19509. .26115
                     /note="assembly_fragment"
     misc_feature    26216. .39655
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    39756. .54129
                     /note="assembly_fragment"
     misc_feature    54230. .64518
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    64619. .79904
                     /note="assembly_fragment"
     misc_feature    80005. .109845
                     /note="assembly_fragment"
     misc_feature    109946. .141041
                     /note="assembly_fragment"
BASE COUNT    44919 a  25512 c  25520 g  43989 t   1101 others
ORIGIN


  Query Match             1.2%;  Score 48.2;  DB 2;  Length 141041;
  Best Local Similarity   46.5%;  Pred. No. 0.17;
  Matches  155;  Conservative    0;  Mismatches  178;  Indels    0;  Gaps    0;

Qy   1418 gtgaaaatgccgctgatccctgtgaggatgatagaagtaccatcccggtcccgatggaag 1477
          | ||  |||||  |   | |       | | |    ||    || ||    |  |   ||
Db  98342 GAGATCATGCCATTACACTCCAGCCTGGTCAACAGAGCGAGATTCCATCTCAAAAAAAAA 98401

Qy   1478 taagcatggatattcctgttagcaaccatacagtgggagaagatgggttaaaatcaagta 1537
           ||  | | | |||    |||| ||  | |  ||     ||||| ||   | | | ||  |
Db  98402 AAAAAAAGTAAATTGGCATTAGGAAAAACATTGTAAAGGAAGAGGGTAAATAGTTAACAA 98461

Qy   1538 agaacaagacaaaacgcaaatactctgatgttgtagatgatggatcatcacttatgaact 1597
          | |  ||  ||||||  ||  | |  ||  |   | | |||     |   ||  ||
Db  98462 ATAGAAAATCAAAACCAAACCAGTTAGACTTAAAAAAAGATATTAAAAAGATTTGAAAAC 98521

Qy   1598 ggctgaatggaaaaaagaaaagaactggaagtgtgcatcacacagttgctcatccagctg 1657
             | |||   | ||||      || | | |||   | ||||| || |  || |
Db  98522 CTGATGGAGAAAATCTGTAAAGCTGATTAAATATTCATTTAATAGTTGTTCCTGAAGATA 98581

Qy   1658 ggaatttgagcaacaaaaaagtgacacccactgcgagtactcagcatgatgatgagaatg 1717
          |  |||  |||||||||| ||  | |   | |   | |    |  || |  ||     |
Db  98582 GACATTGAAGCAACAAAACAGAAAAAATTATTTAAAATTACAATTATCAAAATATTCTGG 98641

Qy   1718 atactgaaaatggtcttgacacaaatatgcata 1750
          | |   ||||     ||      || || ||||
Db  98642 AAAAACAAAAATTGATTCTATTGAAAATTCATA 98674


RESULT   5
AC092925/c
LOCUS       AC092925               141095 bp    DNA     linear   HTG 09-AUG-2001
DEFINITION  Homo sapiens chromosome 3q clone RP11-78O22, WORKING DRAFT
            SEQUENCE, 10 unordered pieces.
ACCESSION   AC092925
VERSION     AC092925.1  GI:15135288
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 141095)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
```

```
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 141095)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-AUG-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: HCXD
                Center clone name: RP11-78O22
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 133822 bases at least Q40
                Consensus quality: 137507 bases at least Q30
                Consensus quality: 138991 bases at least Q20
                Estimated insert size: 139059; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 4.8x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    34364: contig of 34364 bp in length
            *    34365    34464: gap of unknown length
            *    34465    69010: contig of 34546 bp in length
            *    69011    69110: gap of unknown length
            *    69111    84563: contig of 15453 bp in length
            *    84564    84663: gap of unknown length
            *    84664    99240: contig of 14577 bp in length
            *    99241    99340: gap of unknown length
            *    99341   109623: contig of 10283 bp in length
            *   109624   109723: gap of unknown length
            *   109724   123170: contig of 13447 bp in length
            *   123171   123270: gap of unknown length
            *   123271   129353: contig of 6083 bp in length
            *   129354   129453: gap of unknown length
            *   129454   134225: contig of 4772 bp in length
            *   134226   134325: gap of unknown length
            *   134326   138699: contig of 4374 bp in length
            *   138700   138799: gap of unknown length
            *   138800   141095: contig of 2296 bp in length.
FEATURES             Location/Qualifiers
     source          1. .141095
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3q"
                     /clone="RP11-78O22"
BASE COUNT    45088 a  25322 c  25884 g  43887 t    914 others
ORIGIN


  Query Match             1.2%;  Score 48.2;  DB 2;  Length 141095;
  Best Local Similarity   46.5%;  Pred. No. 0.17;
  Matches  155;  Conservative    0;  Mismatches  178;  Indels    0;  Gaps    0;

Qy    1418 gtgaaaatgccgctgatccctgtgaggatgatagaagtaccatcccggtcccgatggaag 1477
           | ||  ||||| |  | | |    | | |   ||   || ||   | |  ||
Db   11830 GAGATCATGCCATTACACTCCAGCCTGGTCAACAGAGCGAGATTCCATCTCAAAAAAAAA 11771

Qy    1478 taagcatggatattcctgttagcaaccatacagtgggagaagatgggttaaaatcaagta 1537
            ||  | | | |||    |||| ||  | |  ||    ||||| ||   | | | ||  |
Db   11770 AAAAAAAGTAAATTGGCATTAGGAAAAACATTGTAAAGGAAGAGGGTAAATAGTTAACAA 11711

Qy    1538 agaacaagacaaaacgcaaatactctgatgttgtagatgatggatcatcacttatgaact 1597
           | |  ||  ||||||  ||  | |  ||  |   | | |||     |    ||  ||
Db   11710 ATAGAAAATCAAAACCAAACCAGTTAGACTTAAAAAAAGATATTAAAAAGATTTGAAAAC 11651

Qy    1598 ggctgaatggaaaaaagaaaagaactggaagtgtgcatcacacagttgctcatccagctg 1657
                | |||   | ||||      || | | |||   | |||||| | |  || |
Db   11650 CTGATGGAGAAAATCTGTAAAGCTGATTAAATATTCATTTAATAGTTGTTCCTGAAGATA 11591

Qy    1658 ggaatttgagcaacaaaaaagtgacacccactgcgagtactcagcatgatgatgagaatg 1717
           |  |||  |||||||||| ||  | |   | |   | |    |  || |  ||     |
Db   11590 GACATTGAAGCAACAAAACAGAAAAAATTATTTAAAATTACAATTATCAAAATATTCTGG 11531

Qy    1718 atactgaaaatggtcttgacacaaatatgcata 1750
           | |   ||||     ||       || || ||||
Db   11530 AAAAACAAAAATTGATTCTATTGAAAATTCATA 11498


RESULT   6
AL591122/c
LOCUS       AL591122              189409 bp    DNA     linear   HTG 18-JAN-2002
DEFINITION  Homo sapiens chromosome 1 clone RP11-415K20, *** SEQUENCING IN
            PROGRESS ***, 3 unordered pieces.
ACCESSION   AL591122
VERSION     AL591122.13  GI:16904444
KEYWORDS    HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Martin,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 12, 2001 this sequence version replaced gi:16304972.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
```

```
          -------------- Project Information
          Center project name: bA415K20
          -------------- Summary Statistics
          Assembly program: XGAP4; version 4.5
          Sequencing vector: M13; M77815; 0% of reads
          Sequencing vector: plasmid; L08752; 99% of reads
          Chemistry: Dye-terminator ABI; 0% of reads
          Chemistry: Dye-terminator Big Dye; 99% of reads
          Chemistry: Dye-primer Big Dye; 0% of reads
          Consensus quality: 187859 bases at least Q40
          Consensus quality: 188342 bases at least Q30
          Consensus quality: 188672 bases at least Q20
          Insert size: 189209; sum-of-contigs
          Insert size: 192947; 1.3% error; agarose-fp
          Quality coverage: 10.90x in Q20 bases; sum-of-contigs Quality
          coverage: 10.95x in Q20 bases; agarose-fp
          --------------.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 3 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1   128065: contig of 128065 bp in length
          *   128066 128165: gap of      100 bp
          *   128166   187247: contig of 59082 bp in length
          *   187248 187347: gap of      100 bp
          *   187348   189409: contig of 2062 bp in length.
FEATURES             Location/Qualifiers
     source          1. .189409
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-415K20"
                     /clone_lib="RPCI-11.2"
     misc_feature    1. .128065
                     /note="assembly_fragment:03248
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:left"
     misc_feature    128166. .187247
                     /note="assembly_fragment:04137
                     fragment_chain:1"
     misc_feature    187348. .189409
                     /note="assembly_fragment:04057"
BASE COUNT    47654 a  46587 c  48574 g  46393 t    201 others
ORIGIN


  Query Match             1.2%;  Score 47.2;  DB 2;  Length 189409;
  Best Local Similarity   57.4%;  Pred. No. 0.32;
  Matches   85;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy     245 agattgttgcagtagatcaggagggagctcgtgttgttgggacgaactgtatgcttgctc 304
           | |  | |||||   | | ||| ||||||| ||||    |||  | || | | |  ||| |
Db 137147 ATAGCGATGCAGCGAAGCTGGATGGAGCTGGTGTACCAGGGGTGGACGGAACGGATGCAC 137088

Qy     305 gtggtggaactggtgctgtagcgccagtgttggagctgacagcgacgcctcgtcaggatg 364
           || |    ||| |||| || |||| | ||||      ||   | |   |   |||||
Db 137087 CAGGGGTGGATGGAGCTGGTGCACCAGGGGTGGACAGAGGAGACGCACTAGGGGCGGATG 137028

Qy     365 cagccgctgaagctggtgtagacgaacc 392
           ||| | || | | || || |||| | |
Db 137027 GAGCTGATGCACCAGGGGTGGACGGAGC 137000


RESULT   7
AC055845/c
LOCUS       AC055845               151076 bp    DNA     linear   HTG 21-JAN-2002
DEFINITION  Homo sapiens chromosome 11 clone RP11-682B13 map 11, *** SEQUENCING
            IN PROGRESS ***, 8 unordered pieces.
ACCESSION   AC055845
VERSION     AC055845.7  GI:18252746
KEYWORDS    HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 151076)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone RP11-682B13
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 151076)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jan 21, 2002 this sequence version replaced gi:18182814.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L9718
                Center clone name: 682_B_13
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    39083: contig of 39083 bp in length
            *    39084 39183: gap of      100 bp
            *    39184    63670: contig of 24487 bp in length
            *    63671 63770: gap of      100 bp
            *    63771    71885: contig of 8115 bp in length
            *    71886 71985: gap of      100 bp
            *    71986    82569: contig of 10584 bp in length
            *    82570 82669: gap of      100 bp
            *    82670   114517: contig of 31848 bp in length
            *   114518 114617: gap of      100 bp
            *   114618   126413: contig of 11796 bp in length
            *   126414 126513: gap of      100 bp
            *   126514   142582: contig of 16069 bp in length
            *   142583 142682: gap of      100 bp
            *   142683   151076: contig of 8394 bp in length.
FEATURES             Location/Qualifiers
```

```
     source          1. .151076
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11"
                     /clone="RP11-682B13"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    41137 a  32981 c  33887 g  42247 t    824 others
ORIGIN


  Query Match             1.2%;  Score 46.2;  DB 2;  Length 151076;
  Best Local Similarity   48.0%;  Pred. No. 0.59;
  Matches  132;  Conservative    0;  Mismatches  143;  Indels    0;  Gaps    0;

Qy    1470 gatggaagtaagcatggatattcctgttagcaaccatacagtgggagaagatgggttaaa 1529
           | | | ||| |||   |||    |  ||     |||  | | | || |  | |  |
Db  134300 GGTTGCAGTGAGCCGAGATTGCACCATTGCACTCCAGCCTGGGTGACAGCAAGACTCTGT 134241

Qy    1530 atcaagtaagaacaagacaaaacgcaaatactctgatgttgtagatgatggatcatcact 1589
            ||||   | || |  | | ||   ||| |     |      |     ||  ||
Db  134240 CTCAAAATAAAAAATAAAATAAAATAAAAATCTGAAAAAAAAATTGATTGCTTCTAGGGA 134181

Qy    1590 tatgaactggctgaatggaaaaaagaaaagaactggaagtgtgcatcacacagttgctca 1649
             |||||  ||  ||| |  | ||| ||||  | ||   | ||    ||| |   |||
Db  134180 AGGAAACTGAATGGCTGGGATCAGGAATAGAATGGAAATGTTTCACTGTACATTGTTTCA 134121

Qy    1650 tccagctgggaatttgagcaacaaaaaagtgacacccactgcgagtactcagcatgatga 1709
           | |   | |   ||||| | |||| || ||   ||| | |   | || ||||  |  |
Db  134120 TACTTTTTGAGGTTTGAACCACAAGAATGTATTACCTATTTGCATTAATCAGGGTTCTCC 134061

Qy    1710 tgagaatgatactgaaaatggtcttgacacaaata 1744
            |||||  | |   || | | | |  | | | |||
Db  134060 AGAGAAACAGAACCAACAGGATATATATAGATATA 134026


RESULT   8
AP003434
LOCUS       AP003434               145120 bp    DNA     linear   PLN 06-DEC-2001
DEFINITION  Oryza sativa genomic DNA, chromosome 1, PAC clone:P0452F10.
ACCESSION   AP003434
VERSION     AP003434.2  GI:17385724
KEYWORDS    .
SOURCE      Oryza sativa (cultivar:Nipponbare) DNA, clone:P0452F10.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 145120)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2001) Takuji Sasaki, National Institute of
            Agrobiological Sciences, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     On Dec 6, 2001 this sequence version replaced gi:13429999.
            Genes were predicted from the integrated results of the following:
            GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as SplicePredictor
            (October 1998 version). The genomic sequence was searched against
            NCBI NonRedundant Protein database, nr
            (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at
            RGP. Protein homologies of the coding regions were searched against
            NCBI NonRedundant Protein database with BLASTP2.0. ESTs represent
            the identified cDNA sequences using BLASTN 2.0 with the
            corresponding DDBJ accession no. and RGP clone ID.
            A gene with identity or significant homology to a protein is
            classified based on the protein name to indicate the homology level
            such as same name, 'putative-' and '-like protein'.  A gene without
            significant homology to any protein but with EST homology (covering
            almost the entire length of partial sequence) is classified as an
            'unknown' protein. A gene predicted with a gene prediction program
            is classified as a 'hypothetical' protein.
            The orientation of the sequence is from T7 to SP6 of the PAC clone.
            This sequence of P0452F10 clone has an overlap with P0485D09 (DDBJ:
            AP001859) at the position 92,933 to 145,120 of 3' end.   Detailed
            information on overlap and assembly quality together with
            annotation of this entry is available at
            http://rgp.dna.affrc.go.jp/GenomeSeq.html.
FEATURES            Location/Qualifiers
     source         1. .145120
                    /organism="Oryza sativa"
                    /cultivar="Nipponbare"
                    /db_xref="taxon:4530"
                    /chromosome="1"
                    /clone="P0452F10"
     misc_feature   join(3049. .3420,4142. .4504,4612. .4878,5066. .5206,
                    5211. .5279)
                    /gene="P0452F10.1"
                    /note="Rice myb-related protein 5
                    includes stop codon(s)"
                    /pseudo
     gene           join(3049. .3420,4142. .4504,4612. .4878,5066. .5206,
                    5211. .5279)
                    /gene="P0452F10.1"
     gene           join(8115. .8396,9324. .9341)
                    /gene="P0452F10.2"
     CDS            join(8115. .8396,9324. .9341)
                    /gene="P0452F10.2"
                    /note="hypothetical protein"
                    /codon_start=1
                    /protein_id="BAB78665.1"
                    /db_xref="GI:17385725"
                    /translation="MPRRTPAAAAVHTTRRLSRLPLFSPSSTWERAEAAAADRAGVAA
                    GAPILVVYSSAVACASARLLLGRHARPCTAAPVYLSVAGSLVMILNPLDQYYYYY"
     gene           complement(join(9838. .9932,10020. .10126,10340. .10380,
                    10459. .10464))
                    /gene="P0452F10.3"
     CDS            complement(join(9838. .9932,10020. .10126,10340. .10380,
                    10459. .10464))
                    /gene="P0452F10.3"
                    /note="hypothetical protein"
                    /codon_start=1
                    /protein_id="BAB78666.1"
                    /db_xref="GI:17385726"
                    /translation="MEFGKRIRQVFGRVTGWSVSKTEDIFVVTVRKYLRTVTTGSNSI
                    NLVVVVDGKSAVRMPPVRAMRAIVLAAACTECLARTRN"
     gene           join(10963. .11280,12207. .12285,12396. .12531,13093. .13288,
                    13475. .13543,13695. .13834,14005. .14127,14970. .15162,
                    16452. .16614,17013. .17056,17279. .17369,18595. .18738,
                    18827. .18866,23823. .23963,24030. .24075)
                    /gene="P0452F10.4"
     CDS            join(10963. .11280,12207. .12285,12396. .12531,13093. .13288,
                    13475. .13543,13695. .13834,14005. .14127,14970. .15162,
                    16452. .16614,17013. .17056,17279. .17369,18595. .18738,
                    18827. .18866,23823. .23963,24030. .24075)
                    /gene="P0452F10.4"
                    /note="contains EST AU064167(E4058)
                    similar to Arabidopsis thaliana hypothetical protein
                    F17N18.140
                    unknown protein"
                    /codon_start=1
                    /protein_id="BAB78667.1"
                    /db_xref="GI:17385727"
                    /translation="MAAAAVVHLSVHGRLRRSPELHARPYHRPSLLRCRAFKQEADNG
                    GEEASSSPPPPTTAEARRRRKGPLYKLKAAIQGLAGSRSAAAEAYGGEYQRAVEKAEE
                    IFFSVATQVGRYVITMMSSGVVLGVGFQLSGGDSQMNTLIWYSWLGGVIIGTMIGANS
                    VLEEHCKAGPRNVVITGSPESVLQTINELEENIQEGLSVAKKKQREILLHAKVVGTSC
                    DVCKPEDVKKLVNFAKDELGSIDIWINNAGTNKGFRPLVNFSDEDISQIVSTNLVGSL
                    LCTREAMNVMQHQQKGGHVFNMDGAGSGGSSTPLTAVYGSTKCGLRQFQASLLKESRR
                    SKVGVHTASPGMVLTDLLLSGSSLRNKQMFNLICELPETVARTLVPRMRVVKGSGKAI
                    NYLTPPRILLALVTAWVRRGRWFDEEGRAVYAAEADRIRNWAESRARFSFTDAMEMYT
                    ENTWVSVFSLSVVCAFIILSSSVARLHRMLQPPIELKPCRVMILLDGEAGRSLHLVQT
                    MYGFLCASHFFNIAIIMYIHATARPVQIIPMWRGVHPEQWLMCLEVNDFFGFLTRLDG
                    WSHGGIHVVMYFISHDKCRREKGHSSRSNPPFTDAWDPPRSNATLTQVTVARNAPAII
                    TPWAPGGTGCFAEKRG"
```

```
gene        join(24964. .25375,26027. .26542,27087. .27173,27278. .27348,
            27728. .27804,27960. .28107,28553. .28714,28954. .29274)
            /gene="P0452F10.5"
CDS         join(24964. .25375,26027. .26542,27087. .27173,27278. .27348,
            27728. .27804,27960. .28107,28553. .28714,28954. .29274)
            /gene="P0452F10.5"
            /codon_start=1
            /product="putative receptor protein kinase PERK1"
            /protein_id="BAB78668.1"
            /db_xref="GI:17385728"
            /translation="MSSPSSPPPANQTATPPPANQTAAPPPASNNSSSPPAPGSLSPP
            SLTPPAAPSPPSTTPSSPAAAAAASSSGLTTPVVVGIVVGGLVVLLLVTILLVCLFKK
            KKRHHHHHPPPPPPPHLLHYYGHPPPPPPPPPPFKGDHYGGVYQNWQQNGPPPPPDHV
            LKKVPSHPSPPPPPAPLNVHSGGSGSNYSGGDNSQPLVSPGAALGFSRCTFTYEDLSA
            ATDGFSDANLLGQGGFGYVHKGVLPNGTEVAVKQLRDGSGQGEREFQAEVEIISRVHH
            KHLVTLVGYCISGGKRLLVYEYVPNNTLELHLHGRGRPTMEWPTRLRIALGAAKGLAY
            LHEDCHPKIIHRDIKSANILLDARFEAKVADFGLAKLTSDNNTHVSTRVMGTFGYLAP
            EYASSGQLTEKSDVFSFGVMLLELITGRRPVRSNQSQMDDSLVDWARPLMMRASDDGN
            YDALVDPRLGQEYNGNEMARMIACAAACVRHSARRRPRMSQVVRALEGDVSLDDLNEG
            VRPGHSRFLGSYNSNEYDTGHYNEDLKKFRKMAFGSGNLESSQQTQPTEFVPNRSVSM
            GDARQIPETEMEMGSLKKDGDSNGTRDSQAS"
gene        join(30710. .30762,30889. .30915,31041. .31140,32690. .32836,
            32922. .33008,33095. .33234,33611. .33710)
            /gene="P0452F10.6"
CDS         join(30710. .30762,30889. .30915,31041. .31140,32690. .32836,
            32922. .33008,33095. .33234,33611. .33710)
            /gene="P0452F10.6"
            /note="contains ESTs AU056253(S20425),AU056254(S20425)"
            /codon_start=1
            /product="RAS-related GTP-binding protein Rab7 family"
            /protein_id="BAB78669.1"
            /db_xref="GI:17385729"
            /translation="MASRRRMLLKVIILGDSGVGKTSLMNQYVNKKFSNQYKATIGAD
            FLTKEVQIDDRLFTLQIWDTAGQERFQSLGVAFYRGADCCVLVYDVNVTKSFERLNSW
            REEFLIQASPSDPENFPFVVLGNKIDVDGGNSRTVSEKKAKAWCASKGNIPYFETSAK
            EGFNVEAAFECIARNAIKNEPEEEIIFLIQLMLEVLEGNSARQAVNARRYGTSDIAS"
gene        join(34853. .35857,36458. .37129)
            /gene="P0452F10.7"
CDS         join(34853. .35857,36458. .37129)
            /gene="P0452F10.7"
            /note="contains EST C98812(E3123)"
            /codon_start=1
            /product="putative cytochrome P-450LXXIA1 (cyp71A1)
            family"
            /protein_id="BAB78670.1"
            /db_xref="GI:17385730"
            /translation="MELSSSLAAVLHSPLFLLAALLLLPVFTLLSFSSAKKPGDGGGW
            RLPLPPSPRGVPFLGHLPLLGSLPHRKLRSMAEAHGPVMLLWFGRVPTVVASSAASAQ
            EAMRARDAAFASRARVRMAERLIYGRDMVFAPYGEFWRQARRVSVLHLLSPRRIASFR
            GVREQEVAALLDRVRRRCGVRGGGETVNLSDLLMSYANGVISRAAFGDGAYGLDGDEG
            GEKLRELFANFEALLGTATVGEFVPWLAWVDKLMGLDAKAARISAELDGLLERVIADH
            RERRRLSQPDGGDGDGDGDENVDHRDFVDVLLDVSEVEEGAGAGEVLLFDTVAIKAII
            LDMIAAATDTTFTTLEWAMAELINHPPVMRKLQCEIRAAVGVPGASGGAEVTEDHLGE
            LRLLRAVVKETLRLHAPVPLLVPRETVEDTELLGYRVPARTRVIINVWAIGRDAAAWG
            DRAEEFVPERWLDGGGEEVEYAAQLGQDFRFVPFGAGRRGCPGAGFAAPSIELALTNL
            LYHFDWELPPHADGAAAATAARLDMGELFGLSMRMKTTLNLVAKPWSSDV"
gene        join(39698. .40690,40829. .40832,42030. .42769)
            /gene="P0452F10.8"
CDS         join(39698. .40690,40829. .40832,42030. .42769)
            /gene="P0452F10.8"
            /note="contains EST C98812(E3123)"
            /codon_start=1
            /product="putative cytochrome P-450LXXIA1 (cyp71A1)
            family"
            /protein_id="BAB78671.1"
            /db_xref="GI:17385731"
            /translation="MELSSLAALLHSPLLLAVLLLVFSWLIVSSTKKRPPPPCGDGGR
            RLPLPPSPPGVPLLGHLPLLGTLPHRKLRSMAEAHGPVMLLRLGRVPAVVASSAAAAE
            EVMRTRDLAFASRPRVRMSERLFYGRDMAFAPYGEFWRQARRVTVLHLLSPRRVLSFR
            GVREQEVAALLDRVRRRCGGGGETVNLSDLLMSYAHGVISRAAFGHGGAHGFDGDEGG
            EKLRKLFADFEGLLGTMTVGEFVPWLAWVDKLTGLDAKVARTSAAMDGLLERVIADHR
            ERRRSRGQAVGDGEADADHRDFVDVMLDVSEAEEGAGAGAGGVLFDTVAIKAVILSKK
            VPRVTCMNERFIEQDMMAAGTDSSFTTTEWVMAELINHPRVMRKLQDEIRAVVGTSSA
            SAAAAATGGGQVTEDHLGELPFLRAVIKEMLRLHAPGPLLLPRETVEDTELLGYRIPA
            RTRVIINVWAIGRDAAAWGDSAEEFVPERWLDGGGGGGVEYAQQLGKDSRFVPFGAGR
            RGCPGAGFAALSVELALANLLYHFDWELPPPAASGIMATTRLDMDELFGLSVRLKADL
            NLVAKPWSPGAS"
gene        join(48011. .48970,51142. .51771)
            /gene="P0452F10.9"
CDS         join(48011. .48970,51142. .51771)
            /gene="P0452F10.9"
            /note="contains EST AU163704(E4276)"
            /codon_start=1
            /product="putative cytochrome P-450LXXIA1 (cyp71A1)
            family"
            /protein_id="BAB78672.1"
            /db_xref="GI:17385732"
            /translation="MAVSLVVVVVVVVIAIVVPLLYLVLLPAWKPARRDDGDGGMRRRL
            PPSPPWGLPLLGHLHLLGALPHRALRSLAAAHGPVLLLRLGRVPVVVVSSAAAAEEVM
            RTRDLEFASRPRVAMAERLLYGGRDVAFAPYGEYWRQTRRICVVHLLSARRVLSFRRV
            REEEAAALVARVRAAGGAVDLVEHLTAYSNTVVSRAVFGDESARGLYGDVDRGRVLRK
            LFDDFVELLGQEPMGELLPWLGWVDALNGMEVKVQRTFEALDGILEKVIDDHRRRRRE
```

```
  Query Match              1.2%;  Score 45.4;  DB 8;  Length 145120;
  Best Local Similarity   73.4%;  Pred. No. 0.98;
  Matches   58;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy     522 ttcaagcccatttctgtagcaaagtttcgacgatgggattgctcgaagtgcttggataa 581
           | |  ||||||||||  |  |||| |||||||  || |||| |||    ||||||| |
Db  134344 TGCCGGCCCATTTTCACTTCCAAACTTTCGACCGTGAGATTCCTCCTCATGCTTGGAAAT 134403

Qy     582 gttgaaaacttcagataat 600
           | |||||| ||||||| ||
Db  134404 GGTGAAAAATTCAGATCAT 134422


RESULT   9
AX083744/c
LOCUS       AX083744               1141 bp    DNA     linear   PAT 28-FEB-2001
DEFINITION  Sequence 22 from Patent WO0111061.
ACCESSION   AX083744
VERSION     AX083744.1  GI:13185472
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 1141)
  AUTHORS   Kunst,L. and Clemens,S.
  TITLE     Regulation of embryonic transcription in plants
  JOURNAL   Patent: WO 0111061-A 22 15-FEB-2001;
            UNIVERSITY OF BRITISH COLUMBIA (CA)
FEATURES             Location/Qualifiers
     source          1. .1141
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
     promoter        1. .1141
                     /note="consensus sequence of A.t., L.a., and B.n. FAE1
                     promoters"
BASE COUNT      123 a     32 c     42 g    112 t    832 others
ORIGIN


  Query Match              1.2%;  Score 45.2;  DB 6;  Length 1141;
  Best Local Similarity   10.8%;  Pred. No. 0.63;
  Matches  109;  Conservative  344;  Mismatches  560;  Indels    0;  Gaps    0;

Qy     404 aatgcgagcatttctccataagagggtatgttgctcttcttcagaagaaggatccaaaat 463
           ::| :  : ||   : ::|   ::   :|   : ::| :   :: |     :|: : | :
Db    1091 WWTNHAKRGATMCWYWYWTGTNRRWCMRTYAMRTWYTRSNANWSCATKBMWWTMKWYATK 1032

Qy     464 tctgctctctatctcggattttccatgaccagaaaaaatgtgatgaacacaaagctagtt 523
           : |  :    |:     :  | :|::     |:: | |:: |: || : ::||  :: :
Db    1031 YRTAWYAMWCAWRNNNMWCATNGYAKSCATNNAMWYATTRWAAYAAAKWARWAGNNMRMY 972

Qy     524 caagcccattttctgtagcaaagtttcgacgatgggattgctcgaagtgcttggataagt 583
            ||    :      |:|:::|:|          |::  : :  :      |::
```

```
Db    971 GAAAGNKWGCMAAMATMGBWWADTAGKMCNNNNNNWTTDVRRMAMKAKNNNNNNAYWTAC 912

Qy    584 tgaaaacttcagataatggaacagcaccaagaactcttcccgcaaagcagaatggcacaa 643
          : :||   ::| ::: :   |: : :  :: : |  |::      |: : : :   : ::
Db    911 YNRAATNNKMATHWMKWTHGAHSKRRTRHHTRTCRRTKYNNNNNNNARTVYWYHHAARRWM 852

Qy    644 gtgatggttgctccatcacatttgttcggagcacttttgtgcctgctagtgttggttccc 703
            :|:          ||: |  :: :   ::|              : ::     ::
Db    851 NAWWTRTNNNNNNNNNNACRNTRTWWABWKHSWCNNNNNNNNNNNNNNNTWCHYTTANABB 792

Qy    704 aaaaagtgtctcctagcacacaatcatctcaagggaagaatgctgatagatcaactcttc 763
            :|       : : |   ::| : | : ::  : :   :  : :|   |     :||:
Db    791 CYRANNNNAAARMARTCNNYMHAAVTTTHTDWCYKTWMNTWYWDMMTTMBTTTTRNMTTS 732

Qy    764 caaagagtgtgcaagaaggcaatgactccaaatgcaatgcgccttctggcaagaatggag 823
            :          |        :::|:: :|      ::           ::::::
Db    731 TNMTNNNNNNMWACTNNNNNNMWKAYYAHATNNWGCWWNNTDARRTNNTTVMRRRWMTNT 672

Qy    824 ctgctgaggccaatactgattcaccaatgaaagatttgcaagggccagcccaaaattatg 883
            |: ::  : :: |  |                    |  ::          : : :  :
Db    671 KTRWYSTTRRHHYTGATNNNNNNNNNNNNNNNNSCCTCTRMMTMRWTMKGDGMTVRKKVK 612

Qy    884 atgtggcagcaaatgtctctgaggacaacacttctgttgatgttggggctttacctgaag 943
            : :|    : ::|:::: :    :|::: :|    : : : :    :   : : :
Db    611 WRDTTCTYVDVWADSWVWWYANWMRCRDVTYTRNNTYCKSYAHSYWYWSNNAMWYRRYSA 552

Qy    944 ttccccagattacatggcacatagaagtaaatggtgcagatcaacctccatccactccaa 1003
              :::|::||:  :  : :: :::|||::: :|    :  :  :  |:::
Db    551 RNWSSMARWTTRNNWWMSGBVRMRWAGTMWWRHWNNNNTDTRYYWWWKRWARBTTTVYDS 492

Qy   1004 aactttctgaagtggtcctcaaaagaaatgaagatgaaaatggaaaaactgaagagactc 1063
          :   : :::   ::   :: ||  :| | :::  :        ::|: :  :|  :: :|
Db    491 MCNAKSMWRGNNWRAMKMWWAANNDAGAMDHWTYWMGNNTMWMRRAWKMMNMAWCRRAYC 432

Qy   1064 ttgttgctgagcagtgcaatttgaccaaagatcctaacccaatgtctggaaaggaacgtg 1123
              : :  :  ::|:: :  ||     ::|::: :|:  : ::    :::: |:
Db    431 CNNNNNRACVWHKHKMWRWTWKYMWKAACNNNNBKAMYMRVAWMMYSRDTTNTDWMMWTS 372

Qy   1124 atcaggttgctgagcagtgcaatttgaccaaagatccgaaaccagtgtctgggcagaaat 1183
          ::::  :|: : :  |        :: |     :: ::::  :  :|  | |:  ::||
Db    371 DWBWHWYTVDYTMMRAWNNNNNNNWRBCKTTSWMWWMDHMNTHCTYGNNTWGSAYBMAAM 312

Qy   1184 gtgagcagatctgcaatgagccatgtgaagaagttgttctcaaaagaagctccaaatcta 1243
          :  |  |: ::|  : ::: :: :         :  :: |  :|:   ::::  |:: :
Db    311 SMWAAGASNBVTYNWCWRMTYMGKTMTNNNNNNKAWYYRTKTVAWCNNRYYYDTAVWTBK 252

Qy   1244 agaggaagacggataagaagttgatgaagaagcagcagcacagcaagaaacgcactgccc 1303
          : :     :  ::    :::::::|::::   ::    :::::: : ::::  :
Db    251 RNYKYCYAYBWYYBMYMGKHHWBWWRRABHRSWNMWWVKCRNKYMVSWHYHAMRYBKWAB 192

Qy   1304 aggctgatgtttcagatgcaaagctttgtcggagaaagccaaaaaaggtgcggcttctat 1363
          |:||   :::   : ::      : :: :   : :: :  ||:   :  : ::::  :  |
Db    191 AVGCNNNWKDRMAHHHWCATNNNMMWWWYAYMHHMHKKGKAAWTNNKTABRDDHBAHVKT 132

Qy   1364 cagaaattataaatgctaaccaggttgaggattctagaagtgacgaagttcat 1416
          :: :: ::: |: :    | ::   : :   |: :::|  ||:  ::  :|::
Db    131 YWYWRYDYWCAMCWMNAKAKVRTAMKHMWYYTDRYVSANNTGVRWMMMRWCMW 79


RESULT  10
AL162614
LOCUS       AL162614               93100 bp    DNA     linear   PRI 17-AUG-2001
DEFINITION  Human DNA sequence from clone RP1-220G17 on chromosome 11p12-14.1,
            complete sequence.
ACCESSION   AL162614
VERSION     AL162614.25  GI:15215907
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 93100)
  AUTHORS   Martin,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 20, 2001 this sequence version replaced gi:15029116.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP1-220G17 is
            from the library RPCI-1 constructed by the group of Pieter de Jong.
            For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2
            IMPORTANT: This sequence is not the entire insert of clone
            RP1-220G17 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RP1-76K20 is at 91101 in this sequence.
            The true right end of clone RP1-296L11 is at 2000 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .93100
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="p12-14.1"
                     /clone="RP1-220G17"
                     /clone_lib="RPCI-1"
BASE COUNT    29127 a  15637 c  16764 g  31572 t
ORIGIN


  Query Match             1.2%;  Score 45.2;  DB 9;  Length 93100;
  Best Local Similarity   52.1%;  Pred. No. 1;
  Matches  101;  Conservative    0;  Mismatches   93;  Indels    0;  Gaps    0;

Qy   2799 tgcgacgcatcttctgagaatgatggatccatcaacattagcaagcttccccaactatgg 2858
          | |  | | | ||   | |   |||    ||||  |   |     | ||    || |||||
Db  66184 TACATCTCTTTTTGCCACATCCATGCTAACATCTATTATTTTTTGATTTTTAAATTATGG 66243

Qy   2859 aacttctagcaggaaccagatggagtctcaacttcataattctcagtatgcacataatca 2918
            |||| ||||||   || ||| ||||||  .  |  |  |  |  | ||   | | || |
Db  66244 CCATTCTTGCAGGAGTAAGGTGGTGTCTCATTGTGGTTCTAATTTGCATTTCCCTGATAA 66303

Qy   2919 gtacaaaggatcaaccagcacatcatatggcagtaacctgaatggaaagattccattgac 2978
           |||  | | | | |     | |||||||   |    |||  || | |  |||| ||||
Db  66304 TTACTGATGTTGAGCATTTTCTTCATATGTTTGCTGGCTGTTTGTATATCTTCCTTTGAG 66363

Qy   2979 attcgaagacttat 2992
          | | |   | ||||
Db  66364 AATTGTCTATTTAT 66377


RESULT  11
AC087825
LOCUS       AC087825               114980 bp    DNA     linear   HTG 24-MAY-2001
DEFINITION  Homo sapiens chromosome 11 clone RP1-220G17 map 11, WORKING DRAFT
            SEQUENCE, 3 unordered pieces.
ACCESSION   AC087825
VERSION     AC087825.2  GI:14192972
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
```

```
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 114980)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone RP1-220G17
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 114980)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
            Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
            Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
            Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
            Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
            Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JAN-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 24, 2001 this sequence version replaced gi:12584323.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L12108
                Center clone name: 220_G_17
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 114348 bases at least Q40
                Consensus quality: 114583 bases at least Q30
                Consensus quality: 114709 bases at least Q20
                Insert size: 115000; agarose-fp
                Insert size: 114780; sum-of-contigs
                Quality coverage: 10.3 in Q20 bases; agarose-fp
                Quality coverage: 10.3 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     5011: contig of 5011 bp in length
            *    5012 5111: gap of      100 bp
            *    5112    73744: contig of 68633 bp in length
            *   73745 73844: gap of      100 bp
            *   73845   114980: contig of 41136 bp in length.
FEATURES            Location/Qualifiers
     source         1. .114980
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="11"
                    /map="11"
                    /clone="RP1-220G17"
                    /clone_lib="RPCI Human PAC library 1"
     misc_feature   1. .5011
                    /note="assembly_fragment
                    clone_end:SP6
                    vector_side:left"
     misc_feature   5112. .73744
                    /note="assembly_fragment"
     misc_feature   73845. .114980
                    /note="assembly_fragment
                    clone_end:T7
                    vector_side:right"
BASE COUNT    37057 a  20292 c  20482 g  36949 t    200 others
ORIGIN


  Query Match             1.2%;  Score 45.2;  DB 2;  Length 114980;
  Best Local Similarity   52.1%;  Pred. No. 1.1;
  Matches  101;  Conservative    0;  Mismatches   93;  Indels    0;  Gaps    0;

Qy    2799 tgcgacgcatcttctgagaatgatggatccatcaacattagcaagcttccccaactatgg 2858
           | |  | | | ||   | |   |||    |||| |   |    | ||   || |||||
Db   74953 TACATCTCTTTTTGCCACATCCATGCTAACATCTATTATTTTTTGATTTTTAAATTATGG 75012

Qy    2859 aacttctagcaggaaccagatggagtctcaacttcataattctcagtatgcacataatca 2918
             |||| ||||||   || ||| ||||||   |  |  |  |  | ||   | | || |
Db   75013 CCATTCTTGCAGGAGTAAGGTGGTGTCTCATTGTGGTTCTAATTTGCATTTCCCTGATAA 75072

Qy    2919 gtacaaaggatcaaccagcacatcatatggcagtaacctgaatggaaagattccattgac 2978
            |||  | | | | |     | |||||||   |    |||  || | |  |||| ||||
Db   75073 TTACTGATGTTGAGCATTTTCTTCATATGTTTGCTGGCTGTTTGTATATCTTCCTTTGAG 75132

Qy    2979 attcgaagacttat 2992
           | | |   | ||||
Db   75133 AATTGTCTATTTAT 75146


RESULT  12
AC022048
LOCUS       AC022048              158392 bp    DNA     linear   HTG 03-MAR-2001
DEFINITION  Homo sapiens clone RP11-15F4, WORKING DRAFT SEQUENCE, 2 ordered
            pieces.
ACCESSION   AC022048
VERSION     AC022048.3  GI:13184125
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 158392)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-15F4
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 158392)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
```

```
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 2, 2001 this sequence version replaced gi:10045377.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L3455
                Center clone name: 15_F_4
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 36% of reads
                Sequencing vector: Plasmid; n/a; 64% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 158197 bases at least Q40
                Consensus quality: 158232 bases at least Q30
                Consensus quality: 158245 bases at least Q20
                Insert size: 163000; agarose-fp
                Insert size: 158292; sum-of-contigs
                Quality coverage: 13.8 in Q20 bases; agarose-fp
                Quality coverage: 14.2 in Q20 b.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1    39274: contig of 39274 bp in length
            *    39275 39374: gap of      100 bp
            *    39375   158392: contig of 119018 bp in length.
FEATURES              Location/Qualifiers
     source           1. .158392
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="RP11-15F4"
                      /clone_lib="RPCI-11 Human Male BAC"
     misc_feature     1. .39274
                      /note="assembly_fragment
                      clone_end:SP6
                      vector_side:left"
     misc_feature     39375. .158392
                      /note="assembly_fragment
                      clone_end:T7
                      vector_side:right"
BASE COUNT    48372 a  27087 c  29475 g  53358 t    100 others
ORIGIN


  Query Match             1.2%;  Score 45.2;  DB 2;  Length 158392;
  Best Local Similarity   52.1%;  Pred. No. 1.1;
  Matches  101;  Conservative    0;  Mismatches   93;  Indels    0;  Gaps    0;

Qy    2799 tgcgacgcatcttctgagaatgatggatccatcaacattagcaagcttccccaactatgg 2858
           | |  | | | ||   | |   |||    |||| |   |     | ||    || |||||
Db   89005 TACATCTCTTTTTGCCACATCCATGCTAACATCTATTATTTTTTGATTTTAAATTATGG 89064

Qy    2859 aacttctagcaggaaccagatggagtctcaacttcataattctcagtatgcacataatca 2918
             |||| ||||||   || ||| ||||||   |  |  |  |  | ||   | | || |
Db   89065 CCATTCTTGCAGGAGTAAGGTGGTGTCTCATTGTGGTTCTAATTTGCATTTCCCTGATAA 89124

Qy    2919 gtacaaaggatcaaccagcacatcatatggcagtaacctgaatggaaagattccattgac 2978
            |||  | | | | |     | |||||||   |    |||  || | |  |||| ||||
Db   89125 TTACTGATGTTGAGCATTTTCTTCATATGTTTGCTGGCTGTTTGTATATCTTCCTTTGAG 89184

Qy    2979 attcgaagacttat 2992
           | | |   | ||||
Db   89185 AATTGTCTATTTAT 89198


RESULT  13
SH1UL45A/c
LOCUS       SH1UL45A                3400 bp    DNA     linear   VRL 28-OCT-1994
DEFINITION  Pseudorabies virus helicase (UL5) and UL4 genes, complete cds.
ACCESSION   L20708
VERSION     L20708.1  GI:431464
KEYWORDS    helicase.
SOURCE      Pseudorabies virus (strain Indiana-Funkhauser) DNA.
  ORGANISM  Pseudorabies virus
            Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
            Alphaherpesvirinae; Varicellovirus.
REFERENCE   1  (bases 1 to 3400)
  AUTHORS   Dean,H.J. and Cheung,A.K.
  TITLE     Identification of the pseudorabies virus UL4 and UL5 (helicase)
            genes
  JOURNAL   Virology 202 (2), 962-967 (1994)
  MEDLINE   94303212
REFERENCE   2  (bases 1 to 3400)
  AUTHORS   Dean,H.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-DEC-1993) Hansi J. Dean, Virology Swine Research
            Unit, U.S. Department of Agriculture, Ames, IA 50010, USA
FEATURES              Location/Qualifiers
     source           1. .3400
                      /organism="Pseudorabies virus"
                      /strain="Indiana-Funkhauser"
                      /db_xref="taxon:10345"
                      /note="The virus was passed in Madin-Darby bovine kidney
                      cells"
     gene             217. .2718
                      /gene="UL5"
     CDS              217. .2718
                      /gene="UL5"
                      /note="putative"
                      /codon_start=1
                      /product="helicase"
                      /protein_id="AAA50986.1"
                      /db_xref="GI:456681"
                      /translation="MAAAAAAEGPRATNATYLNFTSMHGVEPIVERVRELAATPAEAP
                      PPLAWFKSLAAADNPVDIEALELPFAAYLISGNAGSGKSTCIQTLNETMDCVITGSTR
                      VAAQNVYAKLSAAYSSRYVNTIFQEFGFRGNHVQAQLGRYSYACPTSPPTVRELQKRD
                      LVYYWEVLQDISRRVLAGSHEEFARLRALERLTGRAAEHLAFACHGSLPAFARSNVIV
                      IDEAGLLGRHLLTAVVYCWWLINAAYDTTQYAARARPVLVCVGSPTQTDSLESRFEHA
                      RQLCRVRASENLLTYLITNRALREYTDLSRNWAIFINNKRCQEYEFGELMKALEYGLP
                      LTDEHLRLVDSFVVPEAYINNPANLQGWTRLYSSHREVSAYMSRLHAHLKVAGDAQFV
                      VFTLPAYTIVRTAAFDRYREATQQPHLTLDRWLAANAGRITNYSQSRDQDAAALRCEA
                      RAQQGVVLARCEVTYVLNSQVAVTTRLKKLVIGFSGTFEAFAAVLRDDAFVHAQGGSA
                      EYTYRFLSSLLFSGMIAFYNFLQRPGLAPEAVTAAYRRLAAVTAAALRVPEEHFDFSG
                      AAAPAAPAGPGGAADDDDDLFAALSENMLDMLYCHYDFARPETTSEVYAQFLMLKTLF
                      ADRYAALSELFGPAFARAPFETHVDSVSVRGCEVFVGGLRGALLSTALQTDSYTLVGY
                      THATAPAFAEELARRKLHAGTAELLASLDTPRVVLRDQSGFLSILNVNLSDFVESLDD
                      LELDMATYVDYGISSKLAMTIARSQGLSLERVAVCFTRANLRMNSVYVAMSRVISSRF
                      LRMNMNPLREEPERDNGISEHILAALRDGAVHIVY"
     gene             2777. .3214
                      /gene="UL4"
     CDS              2777. .3214
                      /gene="UL4"
                      /function="unknown"
                      /note="putative"
                      /codon_start=1
                      /protein_id="AAA50987.1"
                      /db_xref="GI:431466"
                      /translation="MTALVAYSFYEIKLPGGWAQPGCGQTVCEYERGVRVMATDGCTR
                      CDALAPGRVTIQHGPVLTVLAVDGEPERCSYVFARTWPAAPEGALVMPFSTWSCAERS
                      RRLRGPAGGLLATLVAERALHVTITAYRPDVLRDALREARILE"
BASE COUNT      502 a   1241 c   1160 g    497 t
ORIGIN
```

```
  Query Match             1.1%;  Score 44.4;  DB 14;  Length 3400;
  Best Local Similarity   55.1%;  Pred. No. 1.2;
  Matches   87;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy    242 tggagattgttgcagtagatcaggagggagctcgtgttgttgggacgaactgtatgcttg 301
          || ||    |  ||||||  ||  |||  || | | |        || || ||  ||  |
Db    722 TGCAGCACCTCCCAGTAGTACACCAGGTCGCGCTTCTGCAGCTCGCGCACCGTCGGCGGG 663

Qy    302 ctcgtggtggaactggtgctgtagcgccagtgttggagctgacagcgacgcctcgtcagg 361
          ||||||| | |   |  ||||||||||| | | ||  |||  | |  ||    | |
Db    662 CTCGTGGGGCACGCGTAGCTGTAGCGCCCGAGCTGCGCCTGCACGTGGTTCCCGCGGAAG 603

Qy    362 atgcagccgctgaagctggtgtagacgaaccggcacaa 399
            | |  |   |||| ||||||  ||| | |||| | |
Db    602 CCGAACTCCTGGAAGATGGTGTTCACGTAGCGGCTCGA 565


RESULT  14
AC107868/c
LOCUS       AC107868               65478 bp    DNA     linear   HTG 24-JAN-2002
DEFINITION  Mus musculus clone RP23-405A17, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC107868
VERSION     AC107868.1  GI:18308571
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 65478)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-405A17
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 65478)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
            Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L20687
                Center clone name: 405_A_17
            --------------.
            * NOTE: This record contains 81 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      680: contig of 680 bp in length
            *      681 780: gap of      100 bp
            *      781     1513: contig of 733 bp in length
            *     1514 1613: gap of      100 bp
            *     1614     2327: contig of 714 bp in length
            *     2328 2427: gap of      100 bp
            *     2428     3109: contig of 682 bp in length
            *     3110 3209: gap of      100 bp
            *     3210     3891: contig of 682 bp in length
            *     3892 3991: gap of      100 bp
            *     3992     4701: contig of 710 bp in length
            *     4702 4801: gap of      100 bp
            *     4802     5520: contig of 719 bp in length
            *     5521 5620: gap of      100 bp
            *     5621     6340: contig of 720 bp in length
            *     6341 6440: gap of      100 bp
            *     6441     7172: contig of 732 bp in length
            *     7173 7272: gap of      100 bp
            *     7273     8005: contig of 733 bp in length
            *     8006 8105: gap of      100 bp
            *     8106     8832: contig of 727 bp in length
            *     8833 8932: gap of      100 bp
            *     8933     9642: contig of 710 bp in length
            *     9643 9742: gap of      100 bp
            *     9743    10445: contig of 703 bp in length
            *    10446 10545: gap of      100 bp
            *    10546    11257: contig of 712 bp in length
            *    11258 11357: gap of      100 bp
            *    11358    12073: contig of 716 bp in length
            *    12074 12173: gap of      100 bp
            *    12174    12890: contig of 717 bp in length
            *    12891 12990: gap of      100 bp
            *    12991    13570: contig of 580 bp in length
            *    13571 13670: gap of      100 bp
            *    13671    14394: contig of 724 bp in length
            *    14395 14494: gap of      100 bp
            *    14495    15185: contig of 691 bp in length
            *    15186 15285: gap of      100 bp
            *    15286    16019: contig of 734 bp in length
            *    16020 16119: gap of      100 bp
            *    16120    16761: contig of 642 bp in length
            *    16762 16861: gap of      100 bp
            *    16862    17587: contig of 726 bp in length
            *    17588 17687: gap of      100 bp
            *    17688    18380: contig of 693 bp in length
            *    18381 18480: gap of      100 bp
            *    18481    19181: contig of 701 bp in length
            *    19182 19281: gap of      100 bp
            *    19282    19996: contig of 715 bp in length
            *    19997 20096: gap of      100 bp
            *    20097    20802: contig of 706 bp in length
            *    20803 20902: gap of      100 bp
            *    20903    21636: contig of 734 bp in length
            *    21637 21736: gap of      100 bp
            *    21737    22462: contig of 726 bp in length
            *    22463 22562: gap of      100 bp
            *    22563    23301: contig of 739 bp in length
            *    23302 23401: gap of      100 bp
            *    23402    24081: contig of 680 bp in length
            *    24082 24181: gap of      100 bp
            *    24182    24921: contig of 740 bp in length
            *    24922 25021: gap of      100 bp
            *    25022    25743: contig of 722 bp in length
```

```
*    25744 25843: gap of      100 bp
*    25844    26559: contig of 716 bp in length
*    26560 26659: gap of      100 bp
*    26660    27349: contig of 690 bp in length
*    27350 27449: gap of      100 bp
*    27450    28145: contig of 696 bp in length
*    28146 28245: gap of      100 bp
*    28246    28956: contig of 711 bp in length
*    28957 29056: gap of      100 bp
*    29057    29789: contig of 733 bp in length
*    29790 29889: gap of      100 bp
*    29890    30611: contig of 722 bp in length
*    30612 30711: gap of      100 bp
*    30712    31413: contig of 702 bp in length
*    31414 31513: gap of      100 bp
*    31514    32239: contig of 726 bp in length
*    32240 32339: gap of      100 bp
*    32340    33039: contig of 700 bp in length
*    33040 33139: gap of      100 bp
*    33140    33833: contig of 694 bp in length
*    33834 33933: gap of      100 bp
*    33934    34638: contig of 705 bp in length
*    34639 34738: gap of      100 bp
*    34739    35425: contig of 687 bp in length
*    35426 35525: gap of      100 bp
*    35526    36220: contig of 695 bp in length
*    36221 36320: gap of      100 bp
*    36321    37019: contig of 699 bp in length
*    37020 37119: gap of      100 bp
*    37120    37831: contig of 712 bp in length
*    37832 37931: gap of      100 bp
*    37932    38591: contig of 660 bp in length
*    38592 38691: gap of      100 bp
*    38692    39379: contig of 688 bp in length
*    39380 39479: gap of      100 bp
*    39480    40189: contig of 710 bp in length
*    40190 40289: gap of      100 bp
*    40290    41022: contig of 733 bp in length
*    41023 41122: gap of      100 bp
*    41123    41813: contig of 691 bp in length
*    41814 41913: gap of      100 bp
*    41914    42623: contig of 710 bp in length
*    42624 42723: gap of      100 bp
*    42724    43442: contig of 719 bp in length
*    43443 43542: gap of      100 bp
*    43543    44274: contig of 732 bp in length
*    44275 44374: gap of      100 bp
*    44375    45094: contig of 720 bp in length
*    45095 45194: gap of      100 bp
*    45195    45925: contig of 731 bp in length
*    45926 46025: gap of      100 bp
*    46026    46761: contig of 736 bp in length
*    46762 46861: gap of      100 bp
*    46862    47535: contig of 674 bp in length
*    47536 47635: gap of      100 bp
*    47636    48373: contig of 738 bp in length
*    48374 48473: gap of      100 bp
*    48474    49204: contig of 731 bp in length
*    49205 49304: gap of      100 bp
*    49305    50024: contig of 720 bp in length
*    50025 50124: gap of      100 bp
*    50125    50806: contig of 682 bp in length
*    50807 50906: gap of      100 bp
*    50907    51608: contig of 702 bp in length
*    51609 51708: gap of      100 bp
*    51709    52417: contig of 709 bp in length
*    52418 52517: gap of      100 bp
*    52518    53229: contig of 712 bp in length
*    53230 53329: gap of      100 bp
*    53330    54064: contig of 735 bp in length
*    54065 54164: gap of      100 bp
*    54165    54877: contig of 713 bp in length
*    54878 54977: gap of      100 bp
*    54978    55714: contig of 737 bp in length
*    55715 55814: gap of      100 bp

  Query Match             1.1%;  Score 44.2;  DB 2;  Length 65478;
  Best Local Similarity   48.2%;  Pred. No. 1.9;
  Matches  124;  Conservative    0;  Mismatches  133;  Indels    0;  Gaps    0;

Qy   1041 aaatggaaaaactgaagagactcttgttgctgagcagtgcaatttgaccaaagatcctaa 1100
          |||  |||| |  |||   |     |    |    | | |||    |   ||| | | |
Db  20758 AAAGAGAAATAAAGAAAGAATAAAAGACTTTCTAGAATTCAAGAAAAATGAAGGTACCAT 20699

Qy   1101 cccaatgtctggaaaggaacgtgatcaggttgctgagcagtgcaatttgaccaaagatcc 1160
           |||||  | ||| |   || || | ||       ||| | |     | |  | |||  ||
Db  20698 ACCAATTTATGGGACACAAAGTAAGCACTACTAAGAGGAATATTCATAGCACTAAGTGCC 20639

Qy   1161 gaaaccagtgtctgggcagaaatgtgagcagatctgcaatgagccatgtgaagaagttgt 1220
            |  |     | || | | ||   | |||||  |   |  | |  ||  | |  |
Db  20638 TTCATAAAGAAATTGGAACACATACTAACAGATTTAAGAATACCTTTGAAAGAAGAAGAA 20579

Qy   1221 tctcaaaagaagctccaaatctaagaggaagacggataagaagttgatgaagaagcagca 1280
             || |  |     ||    |||| ||||| | | ||||||  || |||||||| || |
Db  20578 GAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAA 20519

Qy   1281 gcacagcaagaaacgca 1297
          | | |  |||||  | |
Db  20518 GAAGAAAAAGAAGGGGA 20502


RESULT  15
AL133232
LOCUS       AL133232               95097 bp    DNA     linear   PRI 15-FEB-2001
DEFINITION  Human DNA sequence from clone RP5-897D18 on chromosome 20 Contains
            part of a gene for a  putative novel protein, EST, STSs and GSSs,
            complete sequence.
ACCESSION   AL133232
VERSION     AL133232.15  GI:8649008
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 95097)
  AUTHORS   Tracey,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jun 21, 2000 this sequence version replaced gi:8574104.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 20, constructed by the Sanger Centre Chromosome 20
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr20
            This sequence is the entire insert of clone RP5-897D18 The true
            left end of clone RP11-46O3 is at 76409 in this sequence. The true
            right end of clone RP4-539E24 is at 40325 in this sequence. This
            sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. RP5-897D18 is from
```

```
          the library RPCI-5 constructed by the group of Pieter de Jong. For
          further details see
          http://www.chori.org/bacpac/home.htm
          VECTOR: pCYPAC2.
FEATURES             Location/Qualifiers
     source          1. .95097
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="20"
                     /clone="RP5-897D18"
                     /clone_lib="RPCI-5"
     repeat_region   352. .391
                     /note="Alu repeat: matches 1. .40 of consensus"
     repeat_region   398. .703
                     /note="AluSx repeat: matches 1. .304 of consensus"
     repeat_region   713. .878
                     /note="MSTA repeat: matches 1. .191 of consensus"
     repeat_region   1118. .1231
                     /note="MIR repeat: matches 32. .153 of consensus"
     repeat_region   2251. .2436
                     /note="MER58A repeat: matches 1. .60 of consensus"
     repeat_region   2437. .2734
                     /note="AluY repeat: matches 1. .298 of consensus"
     repeat_region   2735. .2754
                     /note="MER58A repeat: matches 60. .216 of consensus"
     repeat_region   2831. .3220
                     /note="L2 repeat: matches 2322. .2710 of consensus"
     repeat_region   3357. .3660
                     /note="AluSq repeat: matches 1. .306 of consensus"
     repeat_region   3708. .3901
                     /note="MER3 repeat: matches 5. .203 of consensus"
     repeat_region   3963. .4284
                     /note="AluSp repeat: matches 1. .310 of consensus"
     repeat_region   4311. .4414
                     /note="MIR repeat: matches 26. .138 of consensus"
     repeat_region   5332. .5406
                     /note="L2 repeat: matches 2617. .2696 of consensus"
     repeat_region   6199. .6576
                     /note="MLT1C repeat: matches 1. .417 of consensus"
     repeat_region   6588. .6734
                     /note="L2 repeat: matches 1990. .2156 of consensus"
     repeat_region   6985. .7285
                     /note="AluJo repeat: matches 1. .296 of consensus"
     repeat_region   7889. .8312
                     /note="MLT1D repeat: matches 12. .497 of consensus"
     repeat_region   8326. .8397
                     /note="MIR repeat: matches 65. .143 of consensus"
     repeat_region   8410. .8697
                     /note="AluJb repeat: matches 1. .302 of consensus"
     repeat_region   8708. .8867
                     /note="MLT1J repeat: matches 48. .209 of consensus"
     repeat_region   8893. .8939
                     /note="MLT1J repeat: matches 269. .313 of consensus"
     repeat_region   9075. .9632
                     /note="279 copies 2 mer cc 62% conserved"
     misc_feature    9448. .9450
                     /note="Tandem repeat. Forced join. Approximately 350 bases
                     missing according to restriction digest data"
     repeat_region   9918. .10017
                     /note="MIR repeat: matches 51. .155 of consensus"
     repeat_region   10443. .10758
                     /note="AluJb repeat: matches 1. .304 of consensus"
     repeat_region   11021. .11313
                     /note="AluSx repeat: matches 1. .296 of consensus"
     repeat_region   11326. .11620
                     /note="AluSx repeat: matches 1. .301 of consensus"
     repeat_region   11625. .11937
                     /note="AluY repeat: matches 1. .311 of consensus"
     repeat_region   11938. .12212
                     /note="L1MD2 repeat: matches 6052. .6326 of consensus"
     repeat_region   12213. .12571
                     /note="MLT1A1 repeat: matches 31. .365 of consensus"
     repeat_region   12617. .12726
                     /note="L1MD2 repeat: matches 5922. .6035 of consensus"
     repeat_region   12727. .13033
                     /note="AluSx repeat: matches 1. .308 of consensus"
     repeat_region   13058. .13458
                     /note="L1MD2 repeat: matches 5502. .5934 of consensus"
     repeat_region   13903. .14269
                     /note="L2 repeat: matches 2323. .2750 of consensus"
     repeat_region   14299. .14778
                     /note="240 copies 2 mer aa 64% conserved"
     repeat_region   16362. .16674
                     /note="AluSq repeat: matches 1. .312 of consensus"
     repeat_region   17161. .17294
                     /note="AluJb repeat: matches 1. .134 of consensus"
     repeat_region   17357. .17436
                     /note="MADE1 repeat: matches 1. .80 of consensus"
     repeat_region   17437. .17620
                     /note="MIR repeat: matches 78. .262 of consensus"
     repeat_region   17734. .17866
                     /note="MIR repeat: matches 20. .154 of consensus"
     misc_feature    17867. .18046
                     /note="match: GSS: Em:AQ077906"
     repeat_region   17944. .17994
                     /note="L2 repeat: matches 2635. .2693 of consensus"
     repeat_region   18047. .18078
                     /note="16 copies 2 mer tt 84% conserved"
     repeat_region   18080. .18393
                     /note="AluY repeat: matches 1. .310 of consensus"
     repeat_region   18394. .18501
                     /note="MER45B repeat: matches 926. .1039 of consensus"
     repeat_region   18585. .18889
                     /note="AluSx repeat: matches 1. .304 of consensus"
     repeat_region   19072. .19149
                     /note="L2 repeat: matches 2628. .2702 of consensus"
     repeat_region   19150. .19567
                     /note="MLT2B repeat: matches 1. .444 of consensus"
     repeat_region   19568. .19949
                     /note="L2 repeat: matches 2233. .2628 of consensus"
     repeat_region   19950. .20251
                     /note="AluSx repeat: matches 1. .302 of consensus"
     repeat_region   20252. .20359
                     /note="L2 repeat: matches 2130. .2233 of consensus"
     repeat_region   21615. .21762
                     /note="MER20 repeat: matches 54. .215 of consensus"
     repeat_region   22190. .22419
                     /note="AluSx repeat: matches 3. .296 of consensus"
     repeat_region   23375. .23685
                     /note="AluY repeat: matches 1. .311 of consensus"
     repeat_region   23794. .24101
                     /note="AluJb repeat: matches 1. .310 of consensus"
     repeat_region   24715. .24954
                     /note="AluSx repeat: matches 3. .237 of consensus"
     repeat_region   25093. .25310
                     /note="AluSg/x repeat: matches 92. .309 of consensus"
     repeat_region   25428. .25732
                     /note="L1MC3 repeat: matches 7451. .7735 of consensus"
     repeat_region   25733. .25995
                     /note="AluSq repeat: matches 40. .293 of consensus"
     repeat_region   25996. .26231
                     /note="L1MC3 repeat: matches 7160. .7451 of consensus"
     repeat_region   26370. .26427
                     /note="AluSc repeat: matches 76. .135 of consensus"
     repeat_region   26428. .26736
                     /note="AluY repeat: matches 1. .310 of consensus"
     repeat_region   26737. .26910
                     /note="AluSc repeat: matches 135. .296 of consensus"
     repeat_region   27259. .27554
                     /note="AluSg repeat: matches 1. .296 of consensus"
     repeat_region   27855. .28078
                     /note="AluY repeat: matches 1. .224 of consensus"
     mRNA            complement(join(<28867. .28921,29503. .>29930))
                     /gene="dJ897D18.1"
                     /note="match: ESTs: Em:AI377332"
                     /evidence=not_experimental
```

```
                     /product="dJ897D18.1 (putative novel transcript)"
     gene            complement(join(28867. .28921,29503. .29930))
                     /gene="dJ897D18.1"
     repeat_region   30614. .30718
                     /note="MIR repeat: matches 73. .179 of consensus"
     repeat_region   30788. .30845
                     /note="29 copies 2 mer at 72% conserved"
     repeat_region   30847. .30941
                     /note="AluJo/FRAM repeat: matches 192. .286 of consensus"
     repeat_region   30946. .31076
                     /note="MIR repeat: matches 106. .240 of consensus"

  Query Match             1.1%;  Score 44.2;  DB 9;  Length 95097;
  Best Local Similarity   44.8%;  Pred. No. 2;
  Matches  169;  Conservative    0;  Mismatches  208;  Indels    0;  Gaps    0;

Qy   1516 gaagatgggttaaaatcaagtaagaacaagacaaaacgcaaatactctgatgttgtagat 1575
          ||||  ||   ||  ||  ||||| |||| | || | | | |    ||      | |
Db  14339 GAAGGAAGGAAGGAAAGAAAAAAGAAAAAGAAAGAAGGGAGAAAGAGAGAAAAAAGAAAA 14398

Qy   1576 gatggatcatcacttatgaactggctgaatggaaaaaagaaaagaactggaagtgtgcat 1635
          ||   |      |   |  || |    |||   ||||||  ||||||  ||||    | |
Db  14399 GAAAAAAGGGAAGGAAGAAAATAAAGGAAAATAAAAAAAGAAAGAAAGGGAAAATGGAAA 14458

Qy   1636 cacacagttgctcatccagctgggaatttgagcaacaaaaaagtgacacccactgcgagt 1695
           | | |        |   ||  |||||     | |||||| ||  || |        ||
Db  14459 GAAAGAAAAATAAAGGAAGGAGGGAAGGAAGGAAACAAAGAAAGGAAAAAAGGAAAGAAA 14518

Qy   1696 actcagcatgatgatgagaatgatactgaaaatggtcttgacacaaatatgcataagaca 1755
             |  |     | ||    |  |  |||        ||| | ||| |    |||| |
Db  14519 GAAAAATAAAGAAACGAAGGAGGGAAGGAAGGAAAGAATGAAAGAAAGAGAAGGAAGAAA 14578

Qy   1756 gatgtctgtcagcatgtatcagaaatctccacacagaggtgctcatcaaaggggaaaaca 1815
          || |      |  |   |  | |||     | | |||       || || | || ||  |
Db  14579 GAAGGAAAGAAAGAAAAAGGAAAAAAGGAAAGAAAGAAAAATAAATGAAGGAGGGAAGGA 14638

Qy   1816 gcgggtttgagtaaggggaaaacacattcagctgctagtaccaaatatggtggtgaaagc 1875
            |      || | ||  |||| | |   |  ||   |    ||| | ||  |  |||
Db  14639 AGGAAAGAAAGAAGGGAAAAAAGAAAAATAAATGAAGGAGGGAAAGAAGGAAGAAAAAAT 14698

Qy   1876 accagaaatggtcagaa 1892
          |  |||||  |  ||||
Db  14699 AAAAGAAAGAGAAAGAA 14715


Search completed: September 16, 2002, 20:36:45
Job time: 15337 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 16:58:13 ; Search time 563.69 Seconds
(without alignments)
11866.626 Million cell updates/sec

Title: US-09-828-068-1
Perfect score: 3896
Sequence: 1 cgcggctgtcggagcaaacg.....gtcaacaccggagaatttac 3896

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	43	1.1	1788	23	ABL06815	Drosophila melanog
c	2	43	1.1	4773	23	ABL06814	Drosophila melanog
	3	42.8	1.1	574	21	AAC44195	Arabidopsis thalia
	4	41.4	1.1	246	23	AAS90679	DNA encoding novel
	5	41.4	1.1	2456	23	AAS75925	DNA encoding novel
c	6	41.4	1.1	2848	22	AAH54196	S. epidermidis gen
	7	40.8	1.0	2167	21	AAZ32967	Human wild-type pr
	8	40.4	1.0	686	21	AAA09318	Human cancer assoc
	9	40.2	1.0	15518	24	ABL34172	Human immune syste
	10	40.2	1.0	15518	24	ABL34624	Human metastasis a
c	11	39.8	1.0	15122	24	ABL32775	Human immune syste
	12	39.6	1.0	3257	22	AAH54703	S. epidermidis gen
c	13	39.4	1.0	843	23	AAS92080	DNA encoding novel
	14	39.2	1.0	2167	21	AAZ38223	Human S184L mutant
	15	39.2	1.0	2522	19	AAV63196	cDNA from clone fp
c	16	39.2	1.0	6065	24	ABL32504	Human immune syste
c	17	39	1.0	833	22	AAI94740	Human neuroblastom
	18	38.8	1.0	567	21	AAA29550	HIV codon altered
	19	38.8	1.0	14041	22	AAH48024	Internal control B
	20	38.6	1.0	261	22	ABA49199	Human breast cell
	21	38.6	1.0	261	22	ABA67114	Human foetal liver
c	22	38.6	1.0	261	22	ABA75585	Human foetal liver
	23	38.6	1.0	261	22	ABA34204	Probe #12670 for g
c	24	38.6	1.0	261	22	ABA40197	Probe #18663 for g
	25	38.6	1.0	261	22	AAK15554	Human brain expres
c	26	38.6	1.0	261	22	AAK24165	Human brain expres
	27	38.6	1.0	261	22	AAK41287	Human bone marrow
c	28	38.6	1.0	261	22	AAK50216	Human bone marrow
	29	38.6	1.0	261	22	AAI22038	Probe #11971 for g
c	30	38.6	1.0	261	22	AAI27301	Probe #17234 for g
	31	38.6	1.0	261	22	AAI47330	Probe #16016 used
c	32	38.6	1.0	261	22	AAI56167	Probe #24853 used
	33	38.6	1.0	261	22	AAI07733	Probe #7724 used t
	34	38.6	1.0	448	22	ABA52290	Human foetal liver
	35	38.6	1.0	448	22	ABA22091	Probe #557 for gen
	36	38.6	1.0	448	22	AAK00565	Human brain expres
	37	38.6	1.0	448	22	AAK26014	Human bone marrow
	38	38.6	1.0	448	22	AAI10643	Probe #576 for gen
	39	38.6	1.0	448	22	AAI31897	Probe #583 used to
	40	38.6	1.0	448	22	AAI00573	Probe #564 used to
c	41	38.6	1.0	513	22	ABA63135	Human foetal liver
c	42	38.6	1.0	513	22	ABA30385	Probe #8851 for ge
c	43	38.6	1.0	513	22	AAK11566	Human brain expres
c	44	38.6	1.0	513	22	AAK37335	Human bone marrow
c	45	38.6	1.0	513	22	AAI18167	Probe #8100 for ge

ALIGNMENTS

RESULT 1
ABL06815/c
ID ABL06815 standard; cDNA; 1788 BP.
XX
AC ABL06815;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 14927.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR P-PSDB; ABB62712.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 14927; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1788 BP; 539 A; 579 C; 432 G; 238 T; 0 other;

Query Match 1.1%; Score 43; DB 23; Length 1788;
Best Local Similarity 47.3%; Pred. No. 0.17;
Matches 130; Conservative 0; Mismatches 145; Indels 0; Gaps 0;

```
Qy      58 gctgattgattgtcttctgtgatatatccagagctcgtgttttgtggtttgtggtttgtg 117
           |||| |    |   | ||||       ||| ||  | | ||  ||||     |  |  |
Db     646 GCTGTTGTTGTTGTTGCTGTTGCTGGCGCAGCGCCTGCGGTTGTTGGTAGTGTGCATAGG 587

Qy     118 gtttgtgcttggattgttgatgtgctaattcgcggcgttacaagatcactgctggattga 177
           | | ||||      || || |||||    |   |  ||| |     |  |    | |  |
Db     586 GATCGTGCAGTTGCTGCTGCTGTGCCCGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGAA 527

Qy     178 tattgagttgtgcctcggctgtgctggctgtgtgttgattctctcctcgtcgtggtgatc 237
             |  | | ||     | | ||   |||||||  |  | ||    | ||   ||||
Db     526 GTTGTTGGTATGTGGGCGTTATGGCCGCTGTGTTGTATTGCTGGTGTTGTATGGGTGGCG 467

Qy     238 gatatggagattgttgcagtagatcaggagggagctcgtgttgttgggacgaactgtatg 297
           |     |||  |||||| |  | |   ||  | | | |||||| ||   |      |
Db     466 GTGGCAGAGTCTGTTGCTGCTGTTGCAGATAGTGTTGGTGTTGGTGTTGCTGCTGCTGCA 407

Qy     298 cttgctcgtggtggaactggtgctgtagcgccagt 332
           |||||  || ||   ||||| |||  ||  | ||
Db     406 GTTGCTGCTGATGTTGCTGGTACTGCTGCTGCTGT 372
```

RESULT 2
ABL06814/c
ID ABL06814 standard; cDNA; 4773 BP.
XX
AC ABL06814;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 14924.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR P-PSDB; ABB62711.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 14924; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 4773 BP; 1400 A; 1213 C; 989 G; 1171 T; 0 other;

Query Match 1.1%; Score 43; DB 23; Length 4773;
Best Local Similarity 47.3%; Pred. No. 0.31;
Matches 130; Conservative 0; Mismatches 145; Indels 0; Gaps 0;

```
Qy      58 gctgattgattgtcttctgtgatatatccagagctcgtgttttgtggtttgtggtttgtg 117
           |||| |    |   | ||||       ||| ||  | | ||  ||||     |  |  |
Db   2965 GCTGTTGTTGTTGTTGCTGTTGCTGGCGCAGCGCCTGCGGTTGTTGGTAGTGTGCATAGG 2906

Qy     118 gtttgtgcttggattgttgatgtgctaattcgcggcgttacaagatcactgctggattga 177
           | | ||||      || || |||||    |   |  ||| |     |  |    | |  |
Db   2905 GATCGTGCAGTTGCTGCTGCTGTGCCCGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGAA 2846

Qy     178 tattgagttgtgcctcggctgtgctggctgtgtgttgattctctcctcgtcgtggtgatc 237
             |  | | ||     | | ||   |||||||  |  | ||    | ||   ||||
Db   2845 GTTGTTGGTATGTGGGCGTTATGGCCGCTGTGTTGTATTGCTGGTGTTGTATGGGTGGCG 2786

Qy     238 gatatggagattgttgcagtagatcaggagggagctcgtgttgttgggacgaactgtatg 297
           |     |||  |||||| |  | |   ||  | | | |||||| ||   |      |
Db   2785 GTGGCAGAGTCTGTTGCTGCTGTTGCAGATAGTGTTGGTGTTGGTGTTGCTGCTGCTGCA 2726

Qy     298 cttgctcgtggtggaactggtgctgtagcgccagt 332
           |||||  || ||   ||||| |||  ||  | ||
Db   2725 GTTGCTGCTGATGTTGCTGGTACTGCTGCTGCTGT 2691
```

RESULT 3
AAC44195
ID AAC44195 standard; DNA; 574 BP.
XX
AC AAC44195;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 41975.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.

PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.1%; Score 42.8; DB 21; Length 574;
Best Local Similarity 57.5%; Pred. No. 0.095;
Matches 77; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

```
Qy   1251 gacggataagaagttgatgaagaagcagcagcacagcaagaaacgcactgcccaggctga 1310
          || |   |||||||||||||||||| || || | |  ||||       ||   | | |||
Db    122 gaagttgaagaagttgatgaagaagaagaagaagaaaaagaggaaatgtgatgatgttga 181

Qy   1311 tgtttcagatgcaaagctttgtcggagaaagccaaaaaaggtgcggcttctatcagaaat 1370
          || |   ||||      ||  ||||| |||    || ||| ||  | ||  | ||  | |
Db    182 tgatgatgatggtggaattgctcggaaaaaaagcaagaagatgaagatttcaacatcagt 241

Qy   1371 tataaatgctaacc 1384
          |  | |   ||| |
Db    242 ttcagacaataatc 255
```

RESULT 4
AAS90679
ID AAS90679 standard; cDNA; 246 BP.
XX
AC AAS90679;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #26483.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG26492.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 26483; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 246 BP; 126 A; 20 C; 77 G; 23 T; 0 other;

Query Match 1.1%; Score 41.4; DB 23; Length 246;
Best Local Similarity 60.0%; Pred. No. 0.14;
Matches 69; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

```
Qy   1178 agaaatgtgagcagatctgcaatgagccatgtgaagaagttgttctcaaaagaagctcca 1237
          || | |||||  |     | ||  ||   || |||||||  |     | |||||     |
Db     71 aggactgtgaagaagaaagaaagaaggactgcgaagaagaagaagaaagaagaaagaaga 130

Qy   1238 aatctaagaggaagacggataagaagttgatgaagaagcagcagcacagcaagaa 1292
          |    |||| ||||| | | ||||||  || |||||||| || || | |  |||||
Db    131 agaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaa 185
```

RESULT 5
AAS75925
ID AAS75925 standard; cDNA; 2456 BP.
XX
AC AAS75925;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #11729.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.

XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG11738.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 11729; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2456 BP; 656 A; 575 C; 654 G; 571 T; 0 other;

Query Match 1.1%; Score 41.4; DB 23; Length 2456;
Best Local Similarity 60.0%; Pred. No. 0.6;
Matches 69; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

Qy 1178 agaaatgtgagcagatctgcaatgagccatgtgaagaagttgttctcaaaagaagctcca 1237
Db 2281 aggactgtgaagaagaaagaaagaaggactgcgaagaagaagaagaaagaagaaagaaga 2340

Qy 1238 aatctaagaggaagacggataagaagttgatgaagaagcagcagcacagcaagaa 1292
Db 2341 agaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaa 2395

RESULT 6
AAH54196/c
ID AAH54196 standard; DNA; 2848 BP.
XX
AC AAH54196;
XX
DT 03-SEP-2001 (first entry)
XX
DE S. epidermidis genomic polynucleotide sequence SEQ ID NO:3560.
XX
KW Staphylococcus epidermidis SR1 strain; infection; diagnosis;
KW vaccination; endocarditis; ds.
XX
OS Staphylococcus epidermidis.
XX
PN WO200134809-A2.
XX
PD 17-MAY-2001.
XX
PF 09-NOV-2000; 2000WO-US30782.
XX
PR 09-NOV-1999; 99US-0164258.
XX
PA (GLAX) GLAXO GROUP LTD.
XX
PI Kimmerly WJ;
XX
DR WPI; 2001-316495/33.
XX
PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PT useful for vaccinating against infections, e.g. endocarditis -
XX
PS Claim 8; Page 1144-1145; 2188pp; English.
XX
CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
CC (I) and (II) can have antibacterial activity and therefore can be used
CC in vaccination. The nucleic acids (I) may be used to produce the
CC S. epidermidis polypeptides (II) via the production of vectors
CC containing them which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to
CC AAH55098 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.
CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO:4454 so even
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC no sequences are present for SEQ ID NO:4455 to 4464.
XX
SQ Sequence 2848 BP; 876 A; 501 C; 442 G; 1029 T; 0 other;

Query Match 1.1%; Score 41.4; DB 22; Length 2848;
Best Local Similarity 49.8%; Pred. No. 0.66;
Matches 105; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

Qy 713 ctcctagcacacaatcatctcaagggaagaatgctgatagatcaactcttccaaagagtg 772
Db 2021 CTATAAGTGAAAAAGCAATCGAACGTAAAACAGGAGCGCGTGGATTACGTTCAATTATAG 1962

Qy 773 tgcaagaaggcaatgactccaaatgcaatgcgccttctggcaagaatggagctgctgagg 832
Db 1961 AAGAAGCATTAATTGACATCATGTATGATGTACCTTCTTCTGAAAATGTAAGTAAAGTAG 1902

Qy 833 ccaatactgattcaccaatgaaagatttgcaagggccagcccaaaattatgatgtggcag 892
Db 1901 TCATTACTGAACAAACAATTAATGAAGAAATTGAGCCTGAATTATATGATGATGAAGGAA 1842

Qy 893 caaatgtctctgaggacaacacttctgttga 923
Db 1841 ATCTAATTAATAAGAACAAGACATCTGCTTA 1811

RESULT 7
AAZ32967
ID AAZ32967 standard; DNA; 2167 BP.
XX
AC AAZ32967;
XX
DT 09-FEB-2000 (first entry)
XX
DE Human wild-type protein kinase MKK4 gene fragment E.
XX
KW MKK4; mitogen activated protein kinase; MAPK; MAPK pathway; mutation;

```
KW   somatic; signal transduction; apoptosis; stress; cytokine; induction;
KW   phosphorylation; Jun kinase; JNK; p38; tumour; suppressor;
KW   loss of heterozygosity; LOH; cancer; detection; diagnosis; prognosis;
KW   breast cancer; pancreatic cancer; colorectal cancer; testicular cancer;
KW   drug screening; gene therapy; protein replacement therapy; mimetic; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   intron          1..247
FT                   /*tag= a
FT   exon            248..367
FT                   /*tag= b
FT                   /note= "MKK4 exon E"
FT   intron          368..2167
FT                   /*tag= c
XX
PN   US5989885-A.
XX
PD   23-NOV-1999.
XX
PF   13-JUN-1997;   97US-0874186.
XX
PR   10-JAN-1997;   97US-0782482.
XX
PA   (MYRI-) MYRIAD GENETICS INC.
XX
PI   Skolnick MH,  Perry WL,  Tavtigian SV,  Teng DH-;
XX
DR   WPI; 2000-022786/02.
XX
PT   Polynucleotides comprising all or a portion of the tumor suppressor
PT   gene MKK4 locus are useful for diagnosis, prognosis and therapy of
PT   human cancers -
XX
PS   Claim 1; Columns 71-74; 63pp; English.
XX
CC   This sequence represents a fragment of the gene encoding a human wild-
CC   type protein kinase, MKK4, which includes exon E. The gene comprises 11
CC   coding exons. MKK4 (also known as JNKK1 and SEK1) may be involved in a
CC   MAPK (mitogen-activated protein kinase) pathway for the signal
CC   transduction of cytokine-induced and stress-induced apoptosis. MKK4 is
CC   also involved in suppressing a variety of tumours. MKK4 is a dual
CC   specific kinase that activates Jun kinases (JNKs) and p38 (a MAPK) but
CC   not extracellular signal-regulated kinases (ERKs) which are a subgroup
CC   of MAPKs. The JNK and p38 MAPKs are activated via dual phosphorylation
CC   on threonine and tyrosine and then go on to activate proteins further
CC   downstream in signal transduction pathways. Tumour suppressor genes such
CC   as MKK4 are deleted at high frequency in certain tumour types. The
CC   deletions often involve loss of a single allele, which is known as loss
CC   of heterozygosity (LOH), and the remaining allele is presumed to be
CC   non-functional, either because of a pre-existing inherited mutation, or
CC   because of a secondary sporadic mutation. Alternatively, the deletion
CC   may involve homozygous deletion of both alleles. LOH events commonly
CC   involve deletions spanning many megabases of DNA, while homozygous
CC   deletions are relatively small in size, probably due to the proximity of
CC   essential genes. Sequences derived from the MKK4 gene can be used to
CC   detect a portion of the MKK4 locus or its expression product in a tissue
CC   sample for the diagnosis and prognosis of human cancer, and can
CC   also be used to diagnose of a predisposition to breast, pancreatic,
CC   colorectal and testicular cancers, as specific MKK4 mutations have
CC   been found in cell lines derived from such tumours. MKK4 oligonucleotides
CC   are useful for the detection of the nucleotide sequence of a particular
CC   MKK4 allele via PCR, and can be used as probes to detect point mutations,
CC   PCR amplification products and mismatches between the MKK4 gene or mRNA.
CC   MKK4 proteins can be used for screening of drugs which can restore MKK4
CC   gene product function for cancer therapy. MKK4 gene therapy, protein
CC   replacement therapy and protein mimetics that reconstitute the function
CC   of the MKK4 protein may be used for therapy of human cancers which result
CC   from a mutation in the MKK4 gene.
XX
SQ   Sequence 2167 BP; 716 A; 334 C; 347 G; 768 T; 2 other;
```

```
  Query Match             1.0%;  Score 40.8;  DB 21;  Length 2167;
  Best Local Similarity   54.7%;  Pred. No. 0.83;
  Matches   81;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy   3610 ttttgttttgccagtccaagaaacgtcctcctgttactttgtagttgtactcatactagt 3669
          | |||| || ||| |  ||| || | |    || || ||   |  | ||   |||  ||
Db    171 tattgtatttccattttaagtaaaggcaaggtgatatttaagatgtataagaataacaga 230

Qy   3670 gcgcttgtttgtacaaggagaaatgtgtaaccttgttgaaaaaaatgtctcccccatttt 3729
               |  | | |   |||   | |||   | ||   || ||   ||||| || | |||
Db    231 tatttgttatttttatagggtgactgttggatctgtatggaactcatgtctacctcgtttg 290

Qy   3730 gtaattaccataaggaggtttatagtgt 3757
          |||  |   | ||  | || ||||||||
Db    291 ataagttttacaaatatgtatatagtgt 318


RESULT   8
AAA09318
ID   AAA09318 standard; DNA; 686 BP.
XX
AC   AAA09318;
XX
DT   10-AUG-2000  (first entry)
XX
DE   Human cancer associated antigen precursor DNA, clone NY-REN-62.
XX
KW   renal cancer; cancer associated antigen precursor; diagnosis;
KW   cytostatic; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   mat_peptide     2..685
FT                   /*tag= a
FT                   /transl_except= (pos:38..43, aa:Gly)
FT                   /transl_except= (pos:59..64, aa:Gly)
FT                   /transl_except= (pos:281..286, aa:Ala)
FT                   /transl_except= (pos:401..406, aa:Lys)
FT                   /transl_except= (pos:620..625, aa:Lys)
XX
PN   WO200020587-A2.
XX
PD   13-APR-2000.
XX
PF   04-OCT-1999;   99WO-US22873.
XX
PR   05-OCT-1998;   98US-0166300.
PR   05-OCT-1998;   98US-0166350.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Obata Y,  Gout I,  Tureci O,  Sahin U,  Pfreundschuh M,  Scanlan MJ;
PI   Stockert E,  Chen Y,  Old LJ,  Jager E,  Knuth A;
XX
DR   WPI; 2000-303774/26.
DR   P-PSDB; AAY92346.
XX
PT   Preventing, diagnosing and/or treating disorders associated with
PT   abnormal expression of human cancer associated antigens
XX
PS   Claim 57; Page 84; 121pp; English.
XX
CC   AAA09310-20 are novel genes isolated by SEREX screening from a renal
CC   cancer cell line 1973/10.4.  The genes encode cancer associated antigen
CC   precursors.  These gene products are useful in methods for preventing,
CC   diagnosing and/or treating disorders, especially cancer, associated with
CC   abnormal expression of human cancer associated antigens.  The method
CC   comprises contacting a sample from a subject with an agent that
CC   specifically binds to the nucleic acid molecule or expression product
CC   (or fragment) complexed with a human leukocyte antigen (HLA) molecule
CC   and determining the interaction between the agent and the nucleic acid
```

```
CC   molecule or the expression product as a determination of the disorder.
XX
SQ   Sequence 686 BP; 299 A; 85 C; 171 G; 126 T; 5 other;


  Query Match             1.0%;  Score 40.4;  DB 21;  Length 686;
  Best Local Similarity   50.3%;  Pred. No. 0.53;
  Matches   98;  Conservative    0;  Mismatches   97;  Indels    0;  Gaps    0;

Qy    1956 ctcggcaaaggtttctccagctgagcatgatatccaaattatgtctgaccttcatgagca 2015
           ||||| | |   ||   |||  ||  | |  |   || ||| ||||||   | |  || |
Db     472 ctcggtagaatcttactcagaagaaaaagcaaaaaaagttaggtctgaatatgaaaagaa 531

Qy    2016 gagtctacccaagaagaaaaagaagcaaaaacttgaagtgacttcgtgaaaaacagaccat 2075
             | | ||| ||| ||| |   ||| | |||| |||   |||   || ||||  | |
Db     532 actccaagccatgaacaaagaactgcagagacttcaagcagctcaaaaagaacatgcaag 591

Qy    2076 gatagatgacatccccatggatattgttgaactgctagctaaaaaccagcatgagaggca 2135
           | |   | | |  |      |   ||   | |   |    |||   ||||| ||   |
Db     592 gttgcttaaaaatcagtctcagtatgaaaagcnattgaagaaattgcagcaggatgtgat 651

Qy    2136 gcttatgactgagac 2150
           |   ||||   | ||
Db     652 ggaaatgaaaaaaac 666


RESULT   9
ABL34172
ID   ABL34172 standard; DNA; 15518 BP.
XX
AC   ABL34172;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 2145.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP07537.
XX
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful
PT   for diagnosis and treatment of diseases associated with abnormal
PT   cytosine methylation  -
XX
PS   Claim 1; SEQ ID NO 2145; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention.
XX
SQ   Sequence 15518 BP; 3047 A; 362 C; 4183 G; 7926 T; 0 other;


  Query Match             1.0%;  Score 40.2;  DB 24;  Length 15518;
  Best Local Similarity   51.6%;  Pred. No. 4.2;
  Matches  116;  Conservative    0;  Mismatches  108;  Indels    1;  Gaps    1;

Qy      60 tgattgattgtcttctgtgatatatccagagctcgtgttttgtggtttgtggtttgtggt 119
           ||  || |||| || |||| | |      | | ||||  |   || |||| |  ||  | |
Db   14825 tgtgtgtttgtgttgtgtgtttttgtaggggttcgttgtagttgttttgggtgtttggtt 14884

Qy     120 ttgtgcttggattgttgatgtgctaattcgcggcgttacaagatcactgctggattgata 179
             ||   ||  |||||  | |  |      |||  ||       |   || | | | |
Db   14885 gagtaggtgtgttgttcgtattgttgagttcgggatttttttttttttggttgttggtag 14944

Qy     180 ttgagttgtgcctcggctgtgctggctgtgtgttgattctctcctcgtcgtggtgatcga 239
           ||| |||  |  |  | |||| ||| |   | || |||   |  | || || | ||| |
Db   14945 ttgggtttagtgt-agttgtgttggtttgattttaatttatttttgtagtagggatagg 15003

Qy     240 tatggagattgttgcagtagatcaggagggagctcgtgttgttgg 284
             |  || || |||   ||| | ||  ||  |   || || | ||
Db   15004 gttttagttttttgttttaggttagttggaggggagtttatggg 15048


RESULT  10
ABL34624
ID   ABL34624 standard; DNA; 15518 BP.
XX
AC   ABL34624;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human metastasis associated gene SEQ ID NO: 177.
XX
KW   Metastasis associated gene; cytostatic; gene therapy; cancer;
KW   cytosine methylation; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200177376-A2.
XX
PD   18-OCT-2001.
XX
PF   06-APR-2001; 2001WO-EP03970.
XX
PR   06-APR-2000; 2000DE-1019058.
PR   07-APR-2000; 2000DE-1019173.
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-010922/01.
XX
PT   New nucleic acid derived from chemically treated metastasis genes,
PT   useful for diagnosis of cancers by analysis of cytosine methylation,
PT   also for treatment  -
XX
PS   Claim 1; SEQ ID NO 177; 23pp + Sequence Listing; English.
XX
CC   The present invention provides a number of human metastasis associated
CC   genes which are modified by cytosine methylation. The sequences can be
CC   used in the diagnosis and treatment of cancer. The present sequence is
CC   one of the genes of the invention.
XX
SQ   Sequence 15518 BP; 3047 A; 362 C; 4183 G; 7926 T; 0 other;
```

```
  Query Match             1.0%;  Score 40.2;  DB 24;  Length 15518;
  Best Local Similarity   51.6%;  Pred. No. 4.2;
  Matches  116;  Conservative    0;  Mismatches  108;  Indels    1;  Gaps    1;

Qy      60 tgattgattgtcttctgtgatatatccagagctcgtgttttgtggtttgtggtttgtggt 119
           ||  || |||| || |||| | |     | | ||||  |   || |||| |  ||  | |
Db   14825 tgtgtgtttgtgttgtgtgtttttgtaggggttcgttgtagttgttttgggtgtttggtt 14884

Qy     120 ttgtgcttggattgttgatgtgctaattcgcggcgttacaagatcactgctggattgata 179
             ||   ||  |||||  | |  |      |||  ||       |   || | | | |
Db   14885 gagtaggtgtgttgttcgtattgttgagttcgggatttttttttttttggttgttggtag 14944

Qy     180 ttgagttgtgcctcggctgtgctggctgtgtgttgattctctcctcgtcgtggtgatcga 239
           ||| |||  |  |  | |||| ||| |   | || |||   |  | || || | ||| |
Db   14945 ttgggtttagtgt-agttgtgttggtttgattttaatttatttttgtagtagggatagg 15003

Qy     240 tatggagattgttgcagtagatcaggagggagctcgtgttgttgg 284
             |  || || |||   ||| | ||  ||  |   || || | ||
Db   15004 gttttagttttttgttttaggttagttggaggggagttttatggg 15048


RESULT  11
ABL32775/c
ID   ABL32775 standard; DNA; 15122 BP.
XX
AC   ABL32775;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 748.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP07537.
XX
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful
PT   for diagnosis and treatment of diseases associated with abnormal
PT   cytosine methylation -
XX
PS   Claim 1; SEQ ID NO 748; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention.
XX
SQ   Sequence 15122 BP; 3782 A; 385 C; 3759 G; 7196 T; 0 other;

  Query Match             1.0%;  Score 39.8;  DB 24;  Length 15122;
  Best Local Similarity   50.8%;  Pred. No. 5.4;
  Matches   95;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy    2003 accttcatgagcagagtctacccaagaagaaaaagaagcaaaaacttgaagtgactcgtg 2062
           | ||  | || || |  |||  |||  | || || ||  |||||  | ||   | | |
Db    6695 AACTAAACGACCAAAACCTAAACAAACAAAATAATAAAAAAAAAAATCAAAAAAATAATA 6636

Qy    2063 aaaaacagaccatgatagatgacatccccatggatattgttgaactgctagctaaaaacc 2122
           || || | ||||  ||| || |  ||  ||   || |     ||    || |  ||  |
Db    6635 AATAAAAAACCAAAATATATAAACTCTACAAACATCTACACAAAAAAATAACATAATCCA 6576

Qy    2123 agcatgagaggcagcttatgactgagactgattgttctgacatcaaccgtattcaatcca 2182
           |  |  |     |    || |||  |||      |   |  ||||| ||  ||||||
Db    6575 ACTAAAACTATAAAAAAATCACTCTAACTACCGTATTAAAACTCAACCATAAAAAATCCA 6516

Qy    2183 agacaac 2189
           |   |||
Db    6515 AATAAAC 6509


RESULT  12
AAH54703
ID   AAH54703 standard; DNA; 3257 BP.
XX
AC   AAH54703;
XX
DT   03-SEP-2001  (first entry)
XX
DE   S. epidermidis genomic polynucleotide sequence SEQ ID NO:4067.
XX
KW   Staphylococcus epidermidis SR1 strain; infection; diagnosis;
KW   vaccination; endocarditis; ds.
XX
OS   Staphylococcus epidermidis.
XX
PN   WO200134809-A2.
XX
PD   17-MAY-2001.
XX
PF   09-NOV-2000; 2000WO-US30782.
XX
PR   09-NOV-1999;   99US-0164258.
XX
PA   (GLAX ) GLAXO GROUP LTD.
XX
PI   Kimmerly WJ;
XX
DR   WPI; 2001-316495/33.
XX
PT   Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PT   useful for vaccinating against infections, e.g. endocarditis -
XX
PS   Claim 8; Page 1752-1753; 2188pp; English.
XX
CC   AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC   (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
CC   (I) and (II) can have antibacterial activity and therefore can be used
CC   in vaccination. The nucleic acids (I) may be used to produce the
CC   S. epidermidis polypeptides (II) via the production of vectors
CC   containing them which are used to produce hosts cells which express the
CC   polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC   used to vaccinate subjects and to raise antibodies against the bacteria.
CC   The polypeptides may also be used to assay for other inhibitors of their
CC   activity and therefore identify compounds that may be used for the
CC   treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC   AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC   polynucleotide sequences from the present invention. AAH55091 to
CC   AAH55098 represent oligonucleotide sequences and primers which are used
CC   in the exemplification of the present invention.
```

```
CC   N.B. The present invention specifically claims all the polynucleotide
CC   sequences given in the sequence listing of the present specification,
CC   however the sequence listing only goes up to SEQ ID NO:4454 so even
CC   though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC   no sequences are present for SEQ ID NO:4455 to 4464.
XX
SQ   Sequence 3257 BP; 1140 A; 462 C; 586 G; 1069 T; 0 other;


  Query Match             1.0%;  Score 39.6;  DB 22:  Length 3257;
  Best Local Similarity   50.5%;  Pred. No. 2.4;
  Matches   96;  Conservative    0;  Mismatches   94;  Indels    0;  Gaps    0;

Qy    734 aagggaagaatgctgatagatcaactcttccaaagagtgtgcaagaaggcaatgactcca 793
          || | || |  |  |   |   |  | | |||  |  |   ||| |   | ||||  ||
Db     22 aacgtaaaacaggagcgcgtggattacgttcaattatagaagaagcattaattgacatca 81

Qy    794 aatgcaatgcgccttctggcaagaatggagctgctgaggccaatactgattcaccaatga 853
            |   |||  ||||||    | |||| |  |   |  | || ||||||   | |||| |
Db     82 tgtatgatgtaccttcttctgaaaatgtaagtaaagtagtcattactgaacaaacaatta 141

Qy    854 aagatttgcaagggccagcccaaaattatgatgtggcagcaaatgtctctgaggacaaca 913
          | ||       | ||| |    | || ||||||  | |       |   | || |||| |
Db    142 atgaagaaattgagcctgaattatatgatgatgaaggaaatctaattaataagaacaaga 201

Qy    914 cttctgttga 923
          | |||| | |
Db    202 catctgctta 211


RESULT  13
AAS92080/c
ID   AAS92080 standard; cDNA; 843 BP.
XX
AC   AAS92080;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #27884.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG27893.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 27884; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 843 BP; 182 A; 212 C; 190 G; 259 T; 0 other;


  Query Match             1.0%;  Score 39.4;  DB 23;  Length 843;
  Best Local Similarity   54.5%;  Pred. No. 1.2;
  Matches   79;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy   1207 tgtgaagaagttgttctcaaaagaagctccaaatctaagaggaagacggataagaagttg 1266
          |||||   ||    |   ||||  |     ||    |||| ||||| | | ||||||  |
Db    252 TGTGATACAGACAGTAGGAAAAACAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 193

Qy   1267 atgaagaagcagcagcacagcaagaaacgcactgcccaggctgatgtttcagatgcaaag 1326
          | ||||||| || || | |  |||||    |  |   |  | ||||||||      |
Db    192 AAGAAGAAGAAGAAGAAGAAGAAGAAGAGAGAAAACACAGATGTTTCTACAAACCAA 133

Qy   1327 ctttgtcggagaaagccaaaaaagg 1351
            | |||      |||||||  || |
Db    132 GTGTGTATCTTCAAGCCAAGCAATG 108


RESULT  14
AAZ38223
ID   AAZ38223 standard; DNA; 2167 BP.
XX
AC   AAZ38223;
XX
DT   09-FEB-2000  (first entry)
XX
DE   Human S184L mutant protein kinase MKK4 gene fragment E.
XX
KW   MKK4; mitogen activated protein kinase; MAPK; MAPK pathway; mutation;
KW   somatic; signal transduction; apoptosis; stress; cytokine; induction;
KW   phosphorylation; Jun kinase; JNK; p38; tumour; suppressor; mutant;
KW   loss of heterozygosity; LOH; cancer; detection; diagnosis; prognosis;
KW   breast cancer; pancreatic cancer; colorectal cancer; testicular cancer;
KW   drug screening; gene therapy; protein replacement therapy; mimetic; ds.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   intron          1..247
FT                   /*tag=  a
FT   exon            248..367
FT                   /*tag=  b
FT                   /note= "MKK4 mutant exon E"
FT   intron          368..2167
FT                   /*tag=  c
XX
PN   US5989885-A.
XX
PD   23-NOV-1999.
XX
PF   13-JUN-1997;   97US-0874186.
XX
PR   10-JAN-1997;   97US-0782482.
```

```
XX
PA   (MYRI-) MYRIAD GENETICS INC.
XX
PI   Skolnick MH,  Perry WL,  Tavtigian SV,  Teng DH-;
XX
DR   WPI; 2000-022786/02.
XX
PT   Polynucleotides comprising all or a portion of the tumor suppressor
PT   gene MKK4 locus are useful for diagnosis, prognosis and therapy of
PT   human cancers  -
XX
PS   Claim 7; Page -; 63pp; English.
XX
CC   This sequence represents a fragment of the gene encoding human S184L
CC   mutant protein kinase, MKK4, which includes exon E. Exon E contains
CC   a base substitution of T for C at base 285, which changes a Ser codon
CC   (TCG) to a Leu codon (TTG). MKK4 (also known as JNKK1 and SEK1) may be
CC   involved in a MAPK (mitogen-activated protein kinase) pathway for the
CC   signal transduction of cytokine-induced and stress-induced apoptosis.
CC   MKK4 is also involved in suppressing a variety of tumours. MKK4 is a dual
CC   specific kinase that activates Jun kinases (JNKs) and p38 (a MAPK) but
CC   not extracellular signal-regulated kinases (ERKs) which are a subgroup
CC   of MAPKs. The JNK and p38 MAPKs are activated via dual phosphorylation
CC   on threonine and tyrosine and then go on to activate proteins further
CC   downstream in signal transduction pathways. Tumour suppressor genes such
CC   as MKK4 are deleted at high frequency in certain tumour types. The
CC   deletions often involve loss of a single allele, which is known as loss
CC   of heterozygosity (LOH), and the remaining allele is presumed to be
CC   non-functional, either because of a pre-existing inherited mutation, or
CC   because of a secondary sporadic mutation. Alternatively, the deletion
CC   may involve homozygous deletion of both alleles. LOH events commonly
CC   involve deletions spanning many megabases of DNA, while homozygous
CC   deletions are relatively small in size, probably due to the proximity of
CC   essential genes. Sequences derived from the MKK4 gene can be used to
CC   detect a portion of the MKK4 locus or its expression product in a tissue
CC   sample for the diagnosis and prognosis of human cancer, and can
CC   also be used to diagnose of a predisposition to breast, pancreatic,
CC   colorectal and testicular cancers, as specific MKK4 mutations have
CC   been found in cell lines derived from such tumours. MKK4 oligonucleotides
CC   are useful for the detection of the nucleotide sequence of a particular
CC   MKK4 allele via PCR, and can be used as probes to detect point mutations,
CC   PCR amplification products and mismatches between the MKK4 gene or mRNA.
CC   MKK4 proteins can be used for screening of drugs which can restore MKK4
CC   gene product function for cancer therapy. MKK4 gene therapy, protein
CC   replacement therapy and protein mimetics that reconstitute the function
CC   of the MKK4 protein may be used for therapy of human cancers which result
CC   from a mutation in the MKK4 gene.
CC   Note: This sequence is not shown in the specification, but is derived
CC   from the MKK4 gene fragment E shown in columns 71-74.
XX
SQ   Sequence 2167 BP; 716 A; 333 C; 347 G; 769 T; 2 other;


  Query Match             1.0%;  Score 39.2;  DB 21;  Length 2167;
  Best Local Similarity   54.1%;  Pred. No. 2.4;
  Matches   80;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy   3610 ttttgttttgccagtccaagaaacgtcctcctgttactttgtagttgtactcatactagt 3669
            | |||| || ||| |  ||| || | |    || || ||   |  | ||   |||  ||
Db    171 tattgtatttccattttaagtaaaggcaaggtgatatttaagatgtataagaataacaga 230

Qy   3670 gcgcttgtttgtacaaggagaaatgtgtaaccttgttgaaaaaaatgtctcccccatttt 3729
               |  | | |   |||   | |||   | ||   || ||   |||||| ||   |||
Db    231 tatttgttattttatagggtgactgttggatctgtatggaactcatgtctaccttgtttg 290

Qy   3730 gtaattaccataaggaggtttatagtgt 3757
           ||| |   | ||  | || ||||||||
Db    291 ataagttttacaaatatgtatatagtgt 318


RESULT  15
AAV63196
ID   AAV63196 standard; cDNA; 2522 BP.
XX
AC   AAV63196;
XX
DT   13-JAN-1999  (first entry)
XX
DE   cDNA from clone fp273_10 which encodes a secreted protein.
XX
KW   Secreted protein; immune stimulating; suppressing;
KW   haematopoiesis regulating activity; tissue growth activity; activin;
KW   inhibin activity; chemotactic; chemokinetic activity; haemostatic;
KW   thrombolytic activity; anti-inflammatory activity; cadherin;
KW   tumour invasion suppressor activity; tumour inhibition activity; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             119..2443
FT                   /*tag=  a
XX
PN   WO9844113-A1.
XX
PD   08-OCT-1998.
XX
PF   27-MAR-1998;   98WO-US06176.
XX
PR   25-MAR-1998;   98US-0823330.
PR   28-MAR-1997;   97US-0823330.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Agostino MJ,  Jacobs K,  Lavallie ER,  McCoy JM,  Merberg D;
PI   Racie LA,  Spaulding V,  Treacy M;
XX
DR   WPI; 1998-542703/46.
DR   P-PSDB; AAW80411.
XX
PT   New isolated polynucleotide(s) and secreted proteins  - are obtained
PT   from human cDNA libraries prepared from adult testes, foetal brain,
PT   adult brain, adult blood and placenta
XX
PS   Claim 31; Pages 86-88; 124pp; English.
XX
CC   The present sequence encodes a secreted protein. The nucleic acid
CC   sequence is isolated from a human adult placenta cDNA library using
CC   probe AAV63206. The polypeptide may have biological activities such has
CC   e.g. nutritional activity, immune stimulating or suppressing activity,
CC   haematopoiesis regulating activity, tissue growth activity,
CC   activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC   activity, cadherin/tumour invasion suppressor activity, tumour
CC   inhibition activity or other activities.
XX
SQ   Sequence 2522 BP; 740 A; 577 C; 637 G; 565 T; 3 other;


  Query Match             1.0%;  Score 39.2;  DB 19;  Length 2522;
  Best Local Similarity   47.4%;  Pred. No. 2.7;
  Matches  110;  Conservative    2;  Mismatches  120;  Indels    0;  Gaps    0;

Qy   1487 atattcctgttagcaaccatacagtgggagaagatgggttaaaatcaagtaagaacaaga 1546
           | || || |  | ||  | || ||  ||  || ||||         | | |   ||
Db   2102 acatcccaggaaccaggcttagaggagggaaacatggacctgttggagggagctacctac 2161

Qy   1547 caaaacgcaaatactctgatgttgtagatgatggatcatcacttatgaactggctgaatg 1606
           ||     || || | :|   || | |   |  | ||||   | |    |   |||| |||
Db   2162 caggtgccagatgccytcgagtgggaacagcagaatcaaggcctggtgagaagctggatg 2221

Qy   1607 gaaaaaagaaaagaactggaagtgtgcatcacacagttgctcatccagctgggaatttga 1666
           ||||||  :|||||   ||  |  | |||  |   | ||||  | |   | ||  || |
Db   2222 gaaaaattmaaagacaaggctggttacatgtctgggatgctggtgcctgtaggggttggg 2281

Qy   1667 gcaacaaaaaagtgacacccactgcgagtactcagcatgatgatgagaatga 1718
            | |   |   |    |    || |||  |||  ||  || | |   ||||
```

Db 2282 atagctggagccttgttcatcttgggagccctctacagcattaaggttatga 2333

Search completed: September 16, 2002, 20:19:54
Job time: 12101 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 16:58:03 ; Search time 126.92 Seconds
(without alignments)
7540.089 Million cell updates/sec

Title: US-09-828-068-1
Perfect score: 3896
Sequence: 1 cgcggctgtcggagcaaacg.....gtcaacaccggagaatttac 3896

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	62.8	1.6	7218	1	US-08-232-463-14	Sequence 14, Appl
	2	40.8	1.0	2167	2	US-08-874-186-40	Sequence 40, Appl
c	3	37.6	1.0	2190	4	US-09-625-188-19	Sequence 19, Appl
c	4	37.4	1.0	5163	3	US-08-700-651-1	Sequence 1, Appli
c	5	37.4	1.0	5163	3	US-08-928-361B-4	Sequence 4, Appli
c	6	37.4	1.0	5318	3	US-08-700-651-2	Sequence 2, Appli
c	7	37.4	1.0	5318	3	US-08-928-361B-3	Sequence 3, Appli
	8	36.4	0.9	2277	1	US-08-676-967-2	Sequence 2, Appli
	9	36.4	0.9	2277	1	US-08-676-974-2	Sequence 2, Appli
	10	36.4	0.9	2277	2	US-09-098-487-2	Sequence 2, Appli
	11	36.2	0.9	1140	4	US-08-510-133A-32	Sequence 32, Appl
	12	36.2	0.9	1140	4	US-08-585-895-32	Sequence 32, Appl
	13	36.2	0.9	1525	2	US-08-824-996-1	Sequence 1, Appli
	14	36.2	0.9	1526	2	US-08-999-811-3	Sequence 3, Appli
	15	36.2	0.9	1526	3	US-09-042-105-3	Sequence 3, Appli
	16	36.2	0.9	1674	2	US-08-999-811-1	Sequence 1, Appli
	17	36.2	0.9	1674	3	US-09-042-105-1	Sequence 1, Appli
	18	36.2	0.9	1674	5	PCT-US96-09001-1	Sequence 1, Appli
	19	36.2	0.9	1997	3	US-08-795-430-7	Sequence 7, Appli
	20	36.2	0.9	1997	4	US-08-510-133A-34	Sequence 34, Appl
	21	35.8	0.9	289	4	US-09-007-005-17	Sequence 17, Appl
	22	35.8	0.9	289	4	US-09-244-796-17	Sequence 17, Appl
	23	35.8	0.9	11283	2	US-08-603-753D-3	Sequence 3, Appli
	24	35.8	0.9	11283	3	US-09-099-753-3	Sequence 3, Appli
	25	35.8	0.9	11283	4	US-08-986-106-3	Sequence 3, Appli
	26	35.8	0.9	11385	2	US-08-639-501-1	Sequence 1, Appli
	27	35.8	0.9	11385	3	US-09-044-946-1	Sequence 1, Appli
	28	35.8	0.9	11385	3	US-09-044-908-1	Sequence 1, Appli
c	29	34.8	0.9	405	2	US-08-299-074A-1	Sequence 1, Appli
c	30	34.8	0.9	405	4	US-09-399-773-1	Sequence 1, Appli
	31	34.8	0.9	1361	4	US-09-232-191-20	Sequence 20, Appl
	32	34.8	0.9	1361	4	US-09-232-200-20	Sequence 20, Appl
	33	34.8	0.9	1361	4	US-09-232-197-20	Sequence 20, Appl
	34	34.8	0.9	1361	4	US-09-232-201-20	Sequence 20, Appl
	35	34.8	0.9	2885	4	US-09-232-200-36	Sequence 36, Appl
	36	34.8	0.9	2885	4	US-09-232-200-56	Sequence 56, Appl
	37	34.8	0.9	2885	4	US-09-232-197-36	Sequence 36, Appl
	38	34.8	0.9	2885	4	US-09-232-197-56	Sequence 56, Appl
	39	34.8	0.9	2885	4	US-09-232-201-36	Sequence 36, Appl
	40	34.8	0.9	2885	4	US-09-232-201-56	Sequence 56, Appl
c	41	34	0.9	16442	3	US-08-781-891-208	Sequence 208, App
c	42	33.8	0.9	1104	3	US-08-454-295-1	Sequence 1, Appli
c	43	33.8	0.9	1104	4	US-09-431-500A-1	Sequence 1, Appli
c	44	33.8	0.9	1360	3	US-08-454-295-2	Sequence 2, Appli
c	45	33.8	0.9	1360	4	US-09-431-500A-2	Sequence 2, Appli

ALIGNMENTS

```
RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
;   APPLICANT: DORNER, F.
;   APPLICANT: SCHEIFLINGER, F.
;   APPLICANT: FALKNER, F. G.
;   TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
;   NUMBER OF SEQUENCES: 52
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Foley & Lardner
;     STREET: 1800 Diagonal Road, Suite 500
;     CITY: Alexandria
;     STATE: VA
;     COUNTRY: USA
;     ZIP: 22313-0299
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/232,463
;     FILING DATE:
;     CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US/07/935,313
;     FILING DATE:
;     APPLICATION NUMBER: EP 91 114 300.6
;     FILING DATE: 26-AUG-1991
;   ATTORNEY/AGENT INFORMATION:
;     NAME: BENT, Stephen A.
;     REGISTRATION NUMBER: 29,768
;     REFERENCE/DOCKET NUMBER: 30472/114 IMMU
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (703)836-9300
;     TELEFAX: (703)683-4109
;     TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 7218 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   IMMEDIATE SOURCE:
;     CLONE: pTZgpt-F1s
US-08-232-463-14
```

```
  Query Match             1.6%;  Score 62.8;  DB 1;  Length 7218;
  Best Local Similarity   7.2%;  Pred. No. 8.1e-09;
  Matches   31;  Conservative  227;  Mismatches  174;  Indels    0;  Gaps    0;

Qy   1182 atgtgagcagatctgcaatgagccatgtgaagaagttgttctcaaaagaagctccaaatc 1241
          ||||  |||   ||| ||| | | || |    |||| |||    | | |
Db   1497 ATGTAGGCATCACTGTAATTACCTATCTATGCAAGTAGTTAAAGAGATAGAAGAATTTGG 1438

Qy   1242 taagaggaagacggataagaagttgatgaagaagcagcagcacagcaagaaacgcactgc 1301
          || :::::::: ::: :::::  :: :::::::: :: :: : :: :::::: : :  :
Db   1437 TACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1378

Qy   1302 ccaggctgatgtttcagatgcaaagctttgtcggagaaagccaaaaaaggtgcggcttct 1361
            :::  :: :    ::: : ::::    :  ::::::::  :::::::: : ::
Db   1377 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1318

Qy   1362 atcagaaattataaatgctaaccaggttgaggattctagaagtgacgaagttcatcgtga 1421
          :  ::::: : ::: :  ::  :::  :::::    ::::: :: ::::   :  : ::
Db   1317 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1258

Qy   1422 aaatgccgctgatccctgtgaggatgatagaagtaccatcccggtcccgatggaagtaag 1481
          ::: :  :  ::     : ::::: :: ::::: :  :    ::    :: ::::: :::
Db   1257 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1198

Qy   1482 catggatattcctgttagcaaccatacagtgggagaagatgggttaaaatcaagtaagaa 1541
           : ::: :     :  :: ::  : : :: ::::::::: :::  ::::  ::: :::::
Db   1197 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1138

Qy   1542 caagacaaaacgcaaatactctgatgttgtagatgatggatcatcacttatgaactggct 1601
           :::: :::: : ::: :    :: :  : ::: :: :::  :  :   : :::  ::
Db   1137 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1078

Qy   1602 gaatggaaaaaa 1613
          ::: ::::::::|
Db   1077 RRRRRRRRRRA 1066


RESULT   2
US-08-874-186-40
; Sequence 40, Application US/08874186
; Patent No. 5989885
;  GENERAL INFORMATION:
;    APPLICANT:  Teng, David H-F.
;    APPLICANT:  Tavtigian, Sean V.
;    APPLICANT:  Perry III, William L.
;    APPLICANT:  Skolnick, Mark H.
;    TITLE OF INVENTION:  SPECIFIC MUTATIONS OF MAP KINASE KINASE
;    TITLE OF INVENTION:  4 (MKK4) IN HUMAN TUMOR CELL LINES IDENTIFY IT AS A TUMOR
;    TITLE OF INVENTION:  SUPPRESSOR IN VARIOUS TYPES OF CANCER
;    NUMBER OF SEQUENCES:  96
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Venable, Baetjer, Howard & Civiletti, LLP
;      STREET:  1201 New York Avenue, N.W., Suite 1000
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  U.S.A.
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/874,186
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/782,482
;      FILING DATE:  10-JAN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Saxe, Stephen A.
;      REGISTRATION NUMBER:  38,609
;      REFERENCE/DOCKET NUMBER:  24884-121392-01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-962-4848
;      TELEFAX:  202-962-8300
;  INFORMATION FOR SEQ ID NO:  40:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2167 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  intron
;      LOCATION:  1..247
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  248..367
;    FEATURE:
;      NAME/KEY:  intron
;      LOCATION:  368..2167
US-08-874-186-40


  Query Match             1.0%;  Score 40.8;  DB 2;  Length 2167;
  Best Local Similarity   54.7%;  Pred. No. 0.031;
  Matches   81;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy   3610 ttttgttttgccagtccaagaaacgtcctcctgttactttgtagttgtactcatactagt 3669
          | |||| || ||| |  ||| || | |    || || ||   |  | ||   |||  ||
Db    171 TATTGTATTTCCATTTTAAGTAAAGGCAAGGTGATATTTAAGATGTATAAGAATAACAGA 230

Qy   3670 gcgcttgtttgtacaaggagaaatgtgtaaccttgttgaaaaaaatgtctcccccatttt 3729
               | | | |   |||   | |||   | ||   || ||   ||||| || | |||
Db    231 TATTTGTTATTTTATAGGGTGACTGTTGGATCTGTATGGAACTCATGTCTACCTCGTTTG 290

Qy   3730 gtaattaccataaggaggtttatagtgt 3757
           ||| |   | ||  | || ||||||||
Db    291 ATAAGTTTTACAAATATGTATATAGTGT 318


RESULT   3
US-09-625-188-19/c
; Sequence 19, Application US/09625188
; Patent No. 6307037
; GENERAL INFORMATION:
;  APPLICANT:  No. 6307037artis AG
;  TITLE OF INVENTION:  Fungal Target Genes and Methods
;  FILE REFERENCE:  PB/5-31285P1
;  CURRENT APPLICATION NUMBER: US/09/625,188
;  CURRENT FILING DATE:  2000-07-21
;  NUMBER OF SEQ ID NOS:  44
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 19
;   LENGTH: 2190
;   TYPE: DNA
;   ORGANISM: Ashbya gossypii
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(2190)
US-09-625-188-19


  Query Match             1.0%;  Score 37.6;  DB 4;  Length 2190;
  Best Local Similarity   51.8%;  Pred. No. 0.33;
  Matches   85;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy    219 tctcctcgtcgtggtgatcgatatggagattgttgcagtagatcaggagggagctcgtgt 278
          | |  | || || ||    | |  ||    || ||| |  | |   |  || |||  ||
Db   1574 TGTTGTTGTGGTTGTTGCTGTTGAGGTTGCTGCTGCTGCTGCTGCTGTTGGTGCTGCTGC 1515

Qy    279 tgttgggacgaactgtatgcttgctcgtggtggaactggtgctgtagcgccagtgttgga 338
          || ||          |  || |||   || ||  |||| |||||| | | | |   || |
```

```
Db   1514 TGGTGCTGTTGTGAATGAGCCTGCGATTGTTGAGACTGCTGCTGTTGTGGCTGCTGTGAA 1455

Qy    339 gctgacagcgacgcctcgtcaggatgcagccgctgaagctggtg 382
          ||| |   |  ||  |   | ||| || |||| || |||||
Db   1454 AGAGACTGTTGCTGCTGCTGGTGCTGCTGCTGCTGGAGGTGGTG 1411


RESULT   4
US-08-700-651-1/c
; Sequence 1, Application US/08700651B
; Patent No. 6015882
; GENERAL INFORMATION:
;  APPLICANT: PETERSEN, CAROLYN
;  APPLICANT: LEECH, JAMES
;  APPLICANT: NELSON, RICHARD, C.
;  APPLICANT: GUT, JIRI
;  TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS
;  TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
;  TITLE OF INVENTION: INFECTIONS
;  FILE REFERENCE: 480.19-4(HV)
;  CURRENT APPLICATION NUMBER: US/08/700,651B
;  CURRENT FILING DATE: 1997-08-14
;  EARLIER APPLICATION NUMBER: 08/415,751
;  EARLIER FILING DATE: 1995-04-03
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 5163
;   TYPE: DNA
;   ORGANISM: Cryptosporidium parvum
US-08-700-651-1


  Query Match             1.0%;  Score 37.4;  DB 3;  Length 5163;
  Best Local Similarity   46.4%;  Pred. No. 0.72;
  Matches  122;  Conservative    0;  Mismatches  141;  Indels    0;  Gaps    0;

Qy      75 tgtgatatatccagagctcgtgttttgtggtttgtggtttgtggtttgtgcttggattgt 134
           |||| ||     ||   | ||  || ||   |||    |||| ||| || |    ||||
Db    1042 TGTGGTAGTAGTAGTAGTAGTTGTTGGTTTCTTGGTAGTTGTAGTTGTTGTTGTTGTT 983

Qy     135 tgatgtgctaattcgcggcgttacaagatcactgctggattgatattgagttgtgcctcg 194
           || |||| |  ||   |  ||           | | | |    | ||||||   |
Db     982 TGTTGTGGTGGTTGTAGTAGTAGTTGTTGTTGTTGTTGTTGTAGTTGTTGTTGT 923

Qy     195 gctgtgctggctgtgtgttgattctctcctcgtcgtggtgatcgatatggagattgttgc 254
           | || | |   ||| | |    |  | || || ||| | |   | |   | ||||
Db     922 TGGTTTCTTGGTAGTTGTGGTTGTGGTTGTGGTTGTTGTGGTTGTAGTTGTTGTGGTTGT 863

Qy     255 agtagatcaggagggagctcgtgttgttgggacgaactgtatgcttgctcgtggtggaac 314
           ||| |     |  |  | |  |||||||          | |  ||| |   | || |
Db     862 AGTTGTGGTTGTAGTTGTTGTGGTTGTTGTTGTTGTGGTTGTAGTTGTTGTGGTTGTAGT 803

Qy     315 tggtgctgtagcgccagtgttgg 337
           ||  | |||||     ||| | |
Db     802 TGTGGTTGTAGTTGTTGTGGTTG 780


RESULT   5
US-08-928-361B-4/c
; Sequence 4, Application US/08928361B
; Patent No. 6071518
;  GENERAL INFORMATION:
;    APPLICANT:  Petersen, Carolyn
;    TITLE OF INVENTION:  PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
;    TITLE OF INVENTION:  THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
;    TITLE OF INVENTION:  FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
;    TITLE OF INVENTION:  SPECIES INFECTIONS
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PETERS, VERNY, JONES & BIKSA
;      STREET:  385 Sherman Avenue, Suite 6
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94306-1840
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/928,361B
;      FILING DATE:  12-SEP-1997
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/026,062
;      FILING DATE:  13-SEP-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Verny, Hana
;      REGISTRATION NUMBER:  30,518
;      REFERENCE/DOCKET NUMBER:  480.76-1(HV)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-324-1677
;      TELEFAX:  650-324-1678
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5163 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-928-361B-4


  Query Match             1.0%;  Score 37.4;  DB 3;  Length 5163;
  Best Local Similarity   46.4%;  Pred. No. 0.72;
  Matches  122;  Conservative    0;  Mismatches  141;  Indels    0;  Gaps    0;

Qy      75 tgtgatatatccagagctcgtgttttgtggtttgtggtttgtggtttgtgcttggattgt 134
           |||| ||     ||   | ||  || ||   |||    |||| ||| || |    ||||
Db    1041 TGTGGTAGTAGTAGTAGTAGTTGTTGGTTTCTTGGTAGTTGTAGTTGTTGTTGTTGTT 982

Qy     135 tgatgtgctaattcgcggcgttacaagatcactgctggattgatattgagttgtgcctcg 194
           || |||| |  ||   |  ||           | | | |    | ||||||   |
Db     981 TGTTGTGGTGGTTGTAGTAGTAGTTGTTGTTGTTGTTGTTGTAGTTGTTGTTGT 922

Qy     195 gctgtgctggctgtgtgttgattctctcctcgtcgtggtgatcgatatggagattgttgc 254
           | || | |   ||| | |    |  | || || ||| | |   | |   | ||||
Db     921 TGGTTTCTTGGTAGTTGTGGTTGTGGTTGTGGTTGTTGTGGTTGTAGTTGTTGTGGTTGT 862

Qy     255 agtagatcaggagggagctcgtgttgttgggacgaactgtatgcttgctcgtggtggaac 314
           ||| |     |  |  | |  |||||||          | |  ||| |   | || |
Db     861 AGTTGTGGTTGTAGTTGTTGTGGTTGTTGTTGTTGTGGTTGTAGTTGTTGTGGTTGTAGT 802

Qy     315 tggtgctgtagcgccagtgttgg 337
           ||  | |||||     ||| | |
Db     801 TGTGGTTGTAGTTGTTGTGGTTG 779


RESULT   6
US-08-700-651-2/c
; Sequence 2, Application US/08700651B
; Patent No. 6015882
; GENERAL INFORMATION:
;  APPLICANT: PETERSEN, CAROLYN
;  APPLICANT: LEECH, JAMES
;  APPLICANT: NELSON, RICHARD, C.
;  APPLICANT: GUT, JIRI
;  TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS
;  TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
;  TITLE OF INVENTION: INFECTIONS
;  FILE REFERENCE: 480.19-4(HV)
;  CURRENT APPLICATION NUMBER: US/08/700,651B
;  CURRENT FILING DATE: 1997-08-14
;  EARLIER APPLICATION NUMBER: 08/415,751
```

```
;  EARLIER FILING DATE: 1995-04-03
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 5318
;   TYPE: DNA
;   ORGANISM: Cryptosporidium parvum
US-08-700-651-2


  Query Match             1.0%;  Score 37.4;  DB 3;  Length 5318;
  Best Local Similarity   46.4%;  Pred. No. 0.74;
  Matches  122;  Conservative    0;  Mismatches  141;  Indels    0;  Gaps    0;

Qy      75 tgtgatatatccagagctcgtgttttgtggtttgtggtttgtggtttgtgcttggattgt 134
           |||| ||     ||   | ||  || ||   |||    |||| |||  || |    ||||
Db    1042 TGTGGTAGTAGTAGTAGTAGTTGTTGGTTTCTTGGTAGTTGTAGTTGTTGTTGTTGTTGT 983

Qy     135 tgatgtgctaattcgcggcgttacaagatcactgctggattgatattgagttgtgcctcg 194
           || |||| |  ||   |  ||           |  | | |     |  ||||||   |
Db     982 TGTTGTGGTGGTTGTAGTAGTAGTTGTTGTTGTTGTTGTTGTAGTTGTTGTTGT 923

Qy     195 gctgtgctggctgtgtgttgattctctcctcgtcgtggtgatcgatatggagattgttgc 254
           | || | |   ||| | |     |  | || || ||| | |   | |   | ||||
Db     922 TGGTTTCTTGGTAGTTGTGGTTGTGGTTGTGGTTGTTGTGGTTGTAGTTGTTGTGGTTGT 863

Qy     255 agtagatcaggagggagctcgtgttgttgggacgaactgtatgcttgctcgtggtggaac 314
           ||| |     |  |  | |  |||||||          | |  ||| |   | || |
Db     862 AGTTGTGGTTGTAGTTGTTGTGGTTGTTGTTGTTGTGGTTGTAGTTGTTGTGGTTGTAGT 803

Qy     315 tggtgctgtagcgccagtgttgg 337
           ||  | |||||     ||| | |
Db     802 TGTGGTTGTAGTTGTTGTGGTTG 780


RESULT   7
US-08-928-361B-3/c
; Sequence 3, Application US/08928361B
; Patent No. 6071518
;  GENERAL INFORMATION:
;    APPLICANT:  Petersen, Carolyn
;    TITLE OF INVENTION:  PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
;    TITLE OF INVENTION:  THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
;    TITLE OF INVENTION:  FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
;    TITLE OF INVENTION:  SPECIES INFECTIONS
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PETERS, VERNY, JONES & BIKSA
;      STREET:  385 Sherman Avenue, Suite 6
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94306-1840
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/928,361B
;      FILING DATE:  12-SEP-1997
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/026,062
;      FILING DATE:  13-SEP-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Verny, Hana
;      REGISTRATION NUMBER:  30,518
;      REFERENCE/DOCKET NUMBER:  480.76-1(HV)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-324-1677
;      TELEFAX:  650-324-1678
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5318 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-928-361B-3


  Query Match             1.0%;  Score 37.4;  DB 3;  Length 5318;
  Best Local Similarity   46.4%;  Pred. No. 0.74;
  Matches  122;  Conservative    0;  Mismatches  141;  Indels    0;  Gaps    0;

Qy      75 tgtgatatatccagagctcgtgttttgtggtttgtggtttgtggtttgtgcttggattgt 134
           |||| ||     ||   | ||  || ||   |||    |||| |||  || |    ||||
Db    1041 TGTGGTAGTAGTAGTAGTAGTTGTTGGTTTCTTGGTAGTTGTAGTTGTTGTTGTTGTTGT 982

Qy     135 tgatgtgctaattcgcggcgttacaagatcactgctggattgatattgagttgtgcctcg 194
           || |||| |  ||   |  ||           |  | | |     |  ||||||   |
Db     981 TGTTGTGGTGGTTGTAGTAGTAGTTGTTGTTGTTGTTGTTGTTGTAGTTGTTGTTGT 922

Qy     195 gctgtgctggctgtgtgttgattctctcctcgtcgtggtgatcgatatggagattgttgc 254
           | || | |   ||| | |     |  | || || ||| | |   | |   | ||||
Db     921 TGGTTTCTTGGTAGTTGTGGTTGTGGTTGTGGTTGTTGTGGTTGTAGTTGTTGTGGTTGT 862

Qy     255 agtagatcaggagggagctcgtgttgttgggacgaactgtatgcttgctcgtggtggaac 314
           ||| |     |  |  | |  |||||||          | |  ||| |   | || |
Db     861 AGTTGTGGTTGTAGTTGTTGTGGTTGTTGTTGTTGTGGTTGTAGTTGTTGTGGTTGTAGT 802

Qy     315 tggtgctgtagcgccagtgttgg 337
           ||  | |||||     ||| | |
Db     801 TGTGGTTGTAGTTGTTGTGGTTG 779


RESULT   8
US-08-676-967-2
; Sequence 2, Application US/08676967
; Patent No. 5747317
;  GENERAL INFORMATION:
;    APPLICANT:  COLLINS, KATHLEEN
;    TITLE OF INVENTION:  Human Telomerase
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Science & Technology Law Group
;      STREET:  268 Bush Street, Suite 3200
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/676,967
;      FILING DATE:
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman Ph.D., Richard A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  UCB96-055
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415)343-4341
;      TELEFAX:  (415)343-4342
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2277 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-676-967-2
```

```
  Query Match             0.9%;  Score 36.4;  DB 1;  Length 2277;
  Best Local Similarity   29.7%;  Pred. No. 0.81;
  Matches  116;  Conservative   66;  Mismatches  206;  Indels    3;  Gaps    2;

Qy    908 acaacacttctgttgatgttggggctttacctgaagttccccagattacatggcacatag 967
          | :||  |  |      :|| ||       :|  |: |   :| :   :      |  |
Db    391 AARACNGTNTTYGCNCARTTYGGNGCNGTNYTNGARGTNAAYATHCCNMGNAARCCNGAY 450

Qy    968 aagtaaatggtgcagatcaacctccatccactccaaaacttttctgaagtggtcctcaaaa 1027
            |:|||       |:    |:   | : |: |:||::| :|  |:         :
Db    451 GGNAARATGMGNGGNTTYGGNTTYGTNCARTTYAARAAYYTNYTNGARGCNGGNAARGCN 510

Qy   1028 gaaatgaagatgaaaatggaaaaactgaagagac-tcttgttgctgagcagtgcaatttg 1086
            | :   ||||| ||| |: |: | ||: | :     || || |   | ||     |
Db    511 YTNAARGGNATGAAYATGAARGARATHAARGGNMGNACNGTNGCNGTNGAYTGGGCNGTN 570

Qy   1087 accaaagatcctaacccaatgtctggaaaggaacgtgatcaggttgctgagcagtgcaat 1146
           | ||:||:    |:  :    |   ::  |   : |     |  |  ||: |:::  |:
Db    571 GCNAARGAYAARTAYAARGAYACNCARWSNGTNWSNGCNATHGGNGARGARAARWSNCAY 630

Qy   1147 ttgaccaaagatccgaaaccagtgtctgggcagaaatgtgaccagatctgcaatgagcca 1206
             ::: ||: |:| : |: : ||      : |:    | ||: |:     :|  ||:  :
Db    631 GARWSNAARCAYCARGARWSNGTNAARAARAARGGNMGNGARGARGARGAYATGGARGAR 690

Qy   1207 tgtgaagaagttgttctcaaaagaagctccaaatctaagaggaagacggataagaagttg 1266
             ||: | | :|   : |  |  |   : |   : |:||:|| |   |  |  || | :|
Db    691 GARGARAAYGAYG--AYGAYGAYGAYGAYGAYGAYGARGARGAYGGNGTNTTYGAYGAYG 748

Qy   1267 atgaagaagcagcagcacagcaagaaacgca 1297
           | || ||:| :| :| : |   :||:::  |
Db    749 ARGAYGARGARGARGARAAYATHGARWSNAA 779
```

RESULT 9
US-08-676-974-2
; Sequence 2, Application US/08676974
; Patent No. 5770422
; GENERAL INFORMATION:
; APPLICANT: COLLINS, KATHLEEN
; TITLE OF INVENTION: Human Telomerase
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Science & Technology Law Group
; STREET: 268 Bush Street, Suite 3200
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/676,974
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman Ph.D., Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UCB96-055
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)343-4341
; TELEFAX: (415)343-4342
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2277 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA

US-08-676-974-2


```
  Query Match             0.9%;  Score 36.4;  DB 1;  Length 2277;
  Best Local Similarity   29.7%;  Pred. No. 0.81;
  Matches  116;  Conservative   66;  Mismatches  206;  Indels    3;  Gaps    2;

Qy    908 acaacacttctgttgatgttggggctttacctgaagttccccagattacatggcacatag 967
          | :||  |  |      :|| ||       :|  |: |   :| :   :      |  |
Db    391 AARACNGTNTTYGCNCARTTYGGNGCNGTNYTNGARGTNAAYATHCCNMGNAARCCNGAY 450

Qy    968 aagtaaatggtgcagatcaacctccatccactccaaaacttttctgaagtggtcctcaaaa 1027
            |:|||       |:    |:   | : |: |:||::| :|  |:         :
Db    451 GGNAARATGMGNGGNTTYGGNTTYGTNCARTTYAARAAYYTNYTNGARGCNGGNAARGCN 510

Qy   1028 gaaatgaagatgaaaatggaaaaactgaagagac-tcttgttgctgagcagtgcaatttg 1086
            | :   ||||| ||| |: |: | ||: | :     || || |   | ||     |
Db    511 YTNAARGGNATGAAYATGAARGARATHAARGGNMGNACNGTNGCNGTNGAYTGGGCNGTN 570

Qy   1087 accaaagatcctaacccaatgtctggaaaggaacgtgatcaggttgctgagcagtgcaat 1146
           | ||:||:    |:  :    |   ::  |   : |     |  |  ||: |:::  |:
Db    571 GCNAARGAYAARTAYAARGAYACNCARWSNGTNWSNGCNATHGGNGARGARAARWSNCAY 630

Qy   1147 ttgaccaaagatccgaaaccagtgtctgggcagaaatgtgagcagatctgcaatgagcca 1206
             ::: ||: |:| : |: : ||      : |:    | ||: |:     :|  ||:  :
Db    631 GARWSNAARCAYCARGARWSNGTNAARAARAARGGNMGNGARGARGARGAYATGGARGAR 690

Qy   1207 tgtgaagaagttgttctcaaaagaagctccaaatctaagaggaagacggataagaagttg 1266
             ||: | | :|   : |  |  |   : |   : |:||:|| |   |  |  || | :|
Db    691 GARGARAAYGAYG--AYGAYGAYGAYGAYGAYGAYGARGARGAYGGNGTNTTYGAYGAYG 748

Qy   1267 atgaagaagcagcagcacagcaagaaacgca 1297
           | || ||:| :| :| : |   :||:::  |
Db    749 ARGAYGARGARGARGARAAYATHGARWSNAA 779
```


RESULT 10
US-09-098-487-2
; Sequence 2, Application US/09098487
; Patent No. 5917025
; GENERAL INFORMATION:
; APPLICANT: COLLINS, Kathleen
; TITLE OF INVENTION: Human Telomerase
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Science & Technology Law Group
; STREET: 268 Bush Street, Suite 3200
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/098,487
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman Ph.D., Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UCB96-055
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)343-4341
; TELEFAX: (415)343-4342
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2277 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear

```
;    MOLECULE TYPE:  cDNA
US-09-098-487-2


  Query Match              0.9%;  Score 36.4;  DB 2;  Length 2277;
  Best Local Similarity   29.7%;  Pred. No. 0.81;
  Matches  116;  Conservative   66;  Mismatches  206;  Indels    3;  Gaps    2;

Qy     908 acaacacttctgttgatgttggggctttacctgaagttccccagattacatggcacatag 967
           | :||  |  |       :|| ||        :|  |: |   :| :   :  :     | |
Db     391 AARACNGTNTTYGCNCARTTYGGNGCNGTNYTNGARGTNAAYATHCCNMGNAARCCNGAY 450

Qy     968 aagtaaatggtgcagatcaacctccatccactccaaaactttctgaagtggtcctcaaaa 1027
             |:|||       |:    |:   | : |: |:||::| :|  |:        :
Db     451 GGNAARATGMGNGGNTTYGGNTTYGTNCARTTYAARAAYYTNYTNGARGCNGGNAARGCN 510

Qy    1028 gaaatgaagatgaaaatggaaaaactgaagagac-tcttgttgctgagcagtgcaatttg 1086
             | :   ||||| ||| |: |: | ||: | :     || || |   | ||     |
Db     511 YTNAARGGNATGAAYATGAARGARATHAARGGNMGNACNGTNGCNGTNGAYTGGGCNGTN 570

Qy    1087 accaaagatcctaacccaatgtctggaaaggaacgtgatcaggttgctgagcagtgcaat 1146
            | ||:||:    |:  :    |   ::  |   : |    |  |  ||: |:::  |:
Db     571 GCNAARGAYAARTAYAARGAYACNCARWSNGTNWSNGCNATHGGNGARGARAARWSNCAY 630

Qy    1147 ttgaccaaagatccgaaaccagtgtctgggcagaaatgtgagcagatctgcaatgagcca 1206
             ::: ||: |:| : |: : ||      : |:    | ||: |:     :|  ||:  :
Db     631 GARWSNAARCAYCARGARWSNGTNAARAARAARGGNMGNGARGARGARGAYATGGARGAR 690

Qy    1207 tgtgaagaagttgttctcaaaagaagctccaaatctaagaggaagacggataagaagttg 1266
              ||: | | :|   : |  |  |   : |   : |:||:|| |   |  |  || | :|
Db     691 GARGARAAYGAYG--AYGAYGAYGAYGAYGAYGAYGARGARGAYGGNGTNTTYGAYGAYG 748

Qy    1267 atgaagaagcagcagcacagcaagaaacgca 1297
             | || ||:| :| :| : |   :||:::  |
Db     749 ARGAYGARGARGARGARAAYATHGARWSNAA 779


RESULT  11
US-08-510-133A-32
; Sequence 32, Application US/08510133A
; Patent No. 6221839
;   GENERAL INFORMATION:
;        APPLICANT: Alitalo, Kari
;                   Joukov, Vladomir
;        TITLE OF INVENTION: Receptor Ligand
;        NUMBER OF SEQUENCES: 35
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
;             STREET: 6300 Sears Tower, 233 South Wacker Drive
;             CITY: Chicago
;             STATE: Illinois
;             COUNTRY: United States of America
;             ZIP: 60606-6402
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.25
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/510,133A
;             FILING DATE: 01-Aug-1995
;             CLASSIFICATION: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Gass, David A.
;             REGISTRATION NUMBER: 38,153
;             REFERENCE/DOCKET NUMBER: 28113/32863
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 312/474-6300
;             TELEFAX: 312/474-0448
;             TELEX: 25-3856
;   INFORMATION FOR SEQ ID NO: 32:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1140 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA (genomic)
;        FEATURE:
;             NAME/KEY:  CDS
;             LOCATION:  37..1089
;        SEQUENCE DESCRIPTION: SEQ ID NO: 32:
US-08-510-133A-32


  Query Match              0.9%;  Score 36.2;  DB 4;  Length 1140;
  Best Local Similarity   53.1%;  Pred. No. 0.56;
  Matches   77;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy    1741 aatatgcataagacagatgtctgtcagcatgtatcagaaatctccacacagaggtgctca 1800
           ||| ||   || ||| ||| | ||  | |||||  ||||    || || | |    | |
Db     827 AATTTGATGAAAACACATGCCAGTGTGTATGTAAAAGAACCTGCCCCAGAAATCAACCCC 886

Qy    1801 tcaaaggggaaaacagcgggtttgagtaaggggaaaacacattcagctgctagtaccaaa 1860
           | ||   | |||  | |  | ||| |    ||||  ||   | |||| ||  ||
Db     887 TAAATCCTGGAAAATGTGCCTGTGAATGTACAGAAAGTCCACAGAAATGCTTGTTAAAAG 946

Qy    1861 tatggtggtgaaagcaccagaaatg 1885
             |  |  ||   | ||||| | |||
Db     947 GAAAGAAGTTCCACCACCAAACATG 971


RESULT  12
US-08-585-895-32
; Sequence 32, Application US/08585895
; Patent No. 6245530
;  GENERAL INFORMATION:
;    APPLICANT:  Alitalo, Kari
;    APPLICANT:  Joukov, Vladimir
;    TITLE OF INVENTION:  Receptor Ligand
;    NUMBER OF SEQUENCES:  35
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/585,895
;      FILING DATE:
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gass, David A.
;      REGISTRATION NUMBER:  38,153
;      REFERENCE/DOCKET NUMBER:  28113/33072
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  32:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1140 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  37..1089
;    FEATURE:
;      NAME/KEY:  mat_peptide
```

```
;      LOCATION:  133..1089
US-08-585-895-32


  Query Match             0.9%;  Score 36.2;  DB 4;  Length 1140;
  Best Local Similarity   53.1%;  Pred. No. 0.56;
  Matches   77;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy   1741 aatatgcataagacagatgtctgtcagcatgtatcagaaatctccacacagaggtgctca 1800
          ||| ||   || ||| ||| | ||  | |||||  ||||    || || | |    | |
Db    827 AATTTGATGAAAACACATGCCAGTGTGTATGTAAAAGAACCTGCCCCAGAAATCAACCCC 886

Qy   1801 tcaaaggggaaaaacagcgggtttgagtaaggggaaaacacattcagctgctagtaccaaa 1860
          | ||    | |||  | |  | ||| |     ||||   ||   |  |||| ||   ||
Db    887 TAAATCCTGGAAAATGTGCCTGTGAATGTACAGAAAGTCCACAGAAATGCTTGTTAAAAG 946

Qy   1861 tatggtggtgaaagcaccagaaatg 1885
           |  |  ||   | ||||| | |||
Db    947 GAAAGAAGTTCCACCACCAAACATG 971


RESULT  13
US-08-824-996-1
; Sequence 1, Application US/08824996B
; Patent No. 5935820
; GENERAL INFORMATION:
;  APPLICANT: Hu, Jing-Shan
;  APPLICANT: Rosen, Craig A.
;  APPLICANT: Cao, Liang
;  TITLE OF INVENTION: Polynucelotides Encoding Vascular Endothelial Growth
;  TITLE OF INVENTION: Factor 2
;  FILE REFERENCE: PF112D1
;  CURRENT APPLICATION NUMBER: US/08/824,996B
;  CURRENT FILING DATE: 1997-03-27
;  EARLIER APPLICATION NUMBER: 08/207,550
;  EARLIER FILING DATE: 1994-03-08
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 1525
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (71)..(1120)
;   FEATURE:
;   NAME/KEY: sig_peptide
;   LOCATION: (71)..(142)
;   FEATURE:
;   NAME/KEY: mat_peptide
;   LOCATION: (143)..(1120)
US-08-824-996-1


  Query Match             0.9%;  Score 36.2;  DB 2;  Length 1525;
  Best Local Similarity   53.1%;  Pred. No. 0.7;
  Matches   77;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy   1741 aatatgcataagacagatgtctgtcagcatgtatcagaaatctccacacagaggtgctca 1800
          ||| ||   || ||| ||| | ||  | |||||  ||||    || || | |    | |
Db    861 aatttgatgaaaacacatgccagtgtgtatgtaaaagaacctgccccagaaatcaacccc 920

Qy   1801 tcaaaggggaaaacagcgggtttgagtaaggggaaaacacattcagctgctagtaccaaa 1860
          | ||    | |||  | |  | ||| |     ||||   ||   |  |||| ||   ||
Db    921 taaatcctggaaaatgtgcctgtgaatgtacagaaagtccacagaaatgcttgttaaaag 980

Qy   1861 tatggtggtgaaagcaccagaaatg 1885
           |  |  ||   | ||||| | |||
Db    981 gaaagaagttccaccaccaaacatg 1005


RESULT  14
US-08-999-811-3
; Sequence 3, Application US/08999811
; Patent No. 5932540
;  GENERAL INFORMATION:
;    APPLICANT:  HU, JING-SHAN
;    APPLICANT:  ROSEN, CRAIG A.
;    APPLICANT:  CAO, LIANG
;    TITLE OF INVENTION:  VASCULAR ENDOTHELIAL GROWTH FACTOR 2
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN & FOX
;      STREET:  1100 NEW YORK AVENUE
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/999,811
;      FILING DATE:  HEREWITH
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/207,550
;      FILING DATE:  8-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/465,968
;      FILING DATE:  06-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MARKOWICZ, KAREN R.
;      REGISTRATION NUMBER:  36,351
;      REFERENCE/DOCKET NUMBER:  1488.1000004
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)371-2600
;      TELEFAX:  (202)371-2540
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1526 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  71..142
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  143..1120
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  71..1120
US-08-999-811-3


  Query Match             0.9%;  Score 36.2;  DB 2;  Length 1526;
  Best Local Similarity   53.1%;  Pred. No. 0.7;
  Matches   77;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy   1741 aatatgcataagacagatgtctgtcagcatgtatcagaaatctccacacagaggtgctca 1800
          ||| ||   || ||| ||| | ||  | |||||  ||||    || || | |    | |
Db    861 AATTTGATGAAAACACATGCCAGTGTGTATGTAAAAGAACCTGCCCCAGAAATCAACCCC 920

Qy   1801 tcaaaggggaaaacagcgggtttgagtaaggggaaaacacattcagctgctagtaccaaa 1860
          | ||    | |||  | |  | ||| |     ||||   ||   |  |||| ||   ||
Db    921 TAAATCCTGGAAAATGTGCCTGTGAATGTACAGAAAGTCCACAGAAATGCTTGTTAAAAG 980

Qy   1861 tatggtggtgaaagcaccagaaatg 1885
           |  |  ||   | ||||| | |||
Db    981 GAAAGAAGTTCCACCACCAAACATG 1005
```

```
RESULT  15
US-09-042-105-3
; Sequence 3, Application US/09042105
; Patent No. 6040157
;  GENERAL INFORMATION:
;    APPLICANT:  HU, JING-SHAN
;    APPLICANT:  ROSEN, CRAIG A.
;    APPLICANT:  CAO, LIANG
;    TITLE OF INVENTION:  VASCULAR ENDOTHELIAL GROWTH FACTOR 2
;    NUMBER OF SEQUENCES:  35
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN & FOX
;      STREET:  1100 NEW YORK AVENUE
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/042,105
;      FILING DATE:  HEREWITH
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/207,550
;      FILING DATE:  8-MAR-1994
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/465,968
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  TO BE ASSIGNED
;      FILING DATE:  24-DEC-1997
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ERIC K. STEFFE
;      REGISTRATION NUMBER:  36,688
;      REFERENCE/DOCKET NUMBER:  1488.1000003/EKS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)371-2600
;      TELEFAX:  (202)371-2540
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1526 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  71..142
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  143..1120
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  71..1120
US-09-042-105-3

  Query Match             0.9%;  Score 36.2;  DB 3;  Length 1526;
  Best Local Similarity   53.1%;  Pred. No. 0.7;
  Matches   77;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy   1741 aatatgcataagacagatgtctgtcagcatgtatcagaaatctccacacagaggtgctca 1800
          ||| ||   || ||| ||| | ||  | |||||  ||||    || || | |    | |
Db    861 AATTTGATGAAAACACATGCCAGTGTGTATGTAAAAGAACCTGCCCCAGAAATCAACCCC 920

Qy   1801 tcaaaggggaaaacagcgggtttgagtaaggggaaaacacattcagctgctagtaccaaa 1860
          | ||    | |||  | |  | ||| |     ||||   ||   |  |||| ||   ||
Db    921 TAAATCCTGGAAAATGTGCCTGTGAATGTACAGAAAGTCCACAGAAATGCTTGTTAAAAG 980

Qy   1861 tatggtggtgaaagcaccagaaatg 1885
          |  |  ||   | |||| | |||
Db    981 GAAAGAAGTTCCACCACCAAACATG 1005


Search completed: September 16, 2002, 20:09:49
Job time: 11506 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 15:26:02 ; Search time 3935.68 Seconds
(without alignments)
13360.877 Million cell updates/sec

Title: US-09-828-068-1
Perfect score: 3896
Sequence: 1 cgcggctgtcggagcaaacg.....gtcaacaccggagaatttac 3896

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	79.8	2.0	321	9	AI691520	AI691520 606020C11
2	61.8	1.6	895	12	CNS0071A	AL066286 Drosophil
c 3	57.8	1.5	1101	12	CNS0039G	AL063921 Drosophil
4	56.2	1.4	922	12	CNS0073W	AL066784 Drosophil
c 5	54	1.4	1101	12	CNS016LW	AL106910 Drosophil
6	53.8	1.4	884	12	CNS006U0	AL065923 Drosophil
c 7	52.8	1.4	939	12	CNS00CNG	AL059400 Drosophil
c 8	51.8	1.3	1101	12	CNS00DG1	AL069971 Drosophil
c 9	50.4	1.3	997	12	CNS005TE	AL060767 Drosophil
10	48.6	1.2	909	12	CNS00JTL	AL076720 Drosophil
11	47.4	1.2	985	12	CNS017TG	AL108478 Drosophil
c 12	46.4	1.2	1101	12	CNS00LO0	AL068607 Drosophil
c 13	46.2	1.2	1101	12	CNS017WI	AL108588 Drosophil
14	46	1.2	706	9	AW684844	AW684844 NF022B10N
c 15	45.6	1.2	748	10	BI176637	BI176637 EST517582
16	45.6	1.2	1101	12	CNS00FXE	AL071370 Drosophil
17	45.6	1.2	1101	12	CNS00LT2	AL078714 Drosophil
c 18	45.2	1.2	490	12	AQ584620	AQ584620 RPCI-11-4
19	45.2	1.2	918	12	AG126399	AG126399 Pan trogl
c 20	45	1.2	937	12	CNS006ST	AL065880 Drosophil
21	44.8	1.1	902	12	CNS006QP	AL065804 Drosophil
c 22	44.6	1.1	450	9	AU060996	AU060996 AU060996
23	44.4	1.1	413	9	AI641313	AI641313 fc13d05.y
c 24	44.2	1.1	1134	12	AG130678	AG130678 Pan trogl
c 25	44	1.1	860	12	CNS018FL	AL109275 Drosophil
c 26	44	1.1	861	12	CNS0075A	AL066834 Drosophil
27	44	1.1	941	12	BH156851	BH156851 ENTQY92TR
28	43.8	1.1	826	10	BF242104	BF242104 601879787
29	43.6	1.1	1101	12	CNS017ZT	AL108707 Drosophil
30	43.4	1.1	754	10	BG845969	BG845969 1024011G0
31	43.4	1.1	1101	12	CNS00GPB	AL072375 Drosophil
c 32	43.2	1.1	1043	12	AG135568	AG135568 Pan trogl
c 33	42.6	1.1	517	12	AG127302	AG127302 Pan trogl
34	42.2	1.1	233	12	AZ267626	AZ267626 RPCI-23-1
35	42.2	1.1	1781	9	AW727582	AW727582 GA__Ea001
36	42	1.1	530	10	C93179	C93179 C93179 Dict
37	42	1.1	624	12	AZ031624	AZ031624 RPCI-23-2
38	42	1.1	684	9	AW187871	AW187871 BNLGHi134
c 39	42	1.1	943	12	CNS06TZ6	AL415192 T3 end of
40	41.8	1.1	513	12	CNS02CMF	AL191328 Tetraodon
c 41	41.8	1.1	640	10	BE427036	BE427036 PSR6610 I
c 42	41.8	1.1	650	10	BE427548	BE427548 PSR7168 I
c 43	41.8	1.1	905	12	AZ550256	AZ550256 ENTEV58TR
c 44	41.8	1.1	1058	10	BE427089	BE427089 PSR6397 I
45	41.6	1.1	819	12	AQ744726	AQ744726 HS_5505_A

ALIGNMENTS

RESULT 1
AI691520/c
LOCUS AI691520 321 bp mRNA linear EST 02-FEB-2000
DEFINITION 606020C11.x1 606 - Ear tissue cDNA library from Schmidt lab Zea mays cDNA, mRNA sequence.
ACCESSION AI691520
VERSION AI691520.1 GI:4966664
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 321)
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford University
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 606020 row: C column: 11.
FEATURES Location/Qualifiers
source 1. .321
/organism="Zea mays"
/cultivar="Ohio43"
/db_xref="taxon:4577"
/clone_lib="606 - Ear tissue cDNA library from Schmidt lab"
/tissue_type="mixed"
/dev_stage="ear length from 0.5 cm - 2.0 cm"
/lab_host="XLOLR (Stratagene)"
/note="Organ: immature ear; Vector: pBK-CMV; Site_1: EcoRI ; Site_2: XhoI; Mixed ear tissue cDNA library from Schmidt lab"
BASE COUNT 78 a 91 c 59 g 91 t 2 others

```
ORIGIN

  Query Match             2.0%;  Score 79.8;  DB 9;  Length 321;
  Best Local Similarity   67.7%;  Pred. No. 1.6e-09;
  Matches  111;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy   3263 ggaacagcatagctcaatcttggaccagaggcaagggtaaaatggttcatcccttggatc 3322
          || |  | |  | |||  ||| ||||||| ||||| || |  |  |||||||| ||||||
Db    303 GGTATGGTACGGGTCAGCCTTCGACCAGTGGCAATGGGAGGACCATTCATCCGTTGGATA 244

Qy   3323 ggtttgtgagacaggatatctgtataactaacaagaacccagctgattttactacaatca 3382
          | | |||||  |||||||||||| | |||| ||  ||  |||| ||||| |||  ||| |
Db    243 AGCTCGTGAGGAAGGATATCTGTGTGACTANCAGAAAAGCAGCCGATTTCACTGTAATTA 184

Qy   3383 gtaacgataacgagtatatggattaccgctgaagcagaaagtgg 3426
          || || | |||||||| |||     ||  |||| ||| |  ||
Db    183 GTGACAAGAACGAGTACATGAGAACCCTTTGAAAGAGAGAAAGG 140


RESULT   2
CNS0071A
LOCUS       CNS0071A                 895 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR14B09 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL066286
VERSION     AL066286.1  GI:4945153
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 895)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .895
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR14B09"
                     /note="end : TET3"
BASE COUNT      124 a     80 c    204 g    179 t    308 others
ORIGIN


  Query Match             1.6%;  Score 61.8;  DB 12;  Length 895;
  Best Local Similarity   21.4%;  Pred. No. 0.00011;
  Matches   63;  Conservative  115;  Mismatches  117;  Indels    0;  Gaps    0;

Qy     33 gttgttctagcgtgtgcagcggctagctgattgattgtcttctgtgatatatccagagct 92
          |::||  | : :||:|  : ||   : ||  :| ::|: ::    : |   |  : : :
Db    597 GKKGTGKTKKTKTGKGKKKGGGTGKKGTGGGKGGKKGKTKKTGTGKKTGGKTKKTKKKKK 656

Qy     93 cgtgttttgtggtttgtggtttgtggtttgtgcttggattgttgatgtgctaattcgcgg 152
          :::::::||| :  ::|: :||:::::::| :|:: ||:||  |:|: |  :: | :|
Db    657 TKKKKKKKGTGTKGGKKGKGKTGKKKKKKKKGTKTKKTTTKTTTTTKTKTTGKKKTGGKG 716

Qy    153 cgttacaagatcactgctggattgatattgagttgtgcctcggctgtgctggctgtgtgt 212
          |:     | :  :|  |:  :: : :|:  :||:|  : :  :::  : : :||:::
Db    717 GGKGKGTKGKKKGTKGGGGKGGKKKKTKTKKTKTGKGKKKTKTTKKKTGKTKKKGTKKKK 776

Qy    213 tgattctctcctcgtcgtggtgatcgatatggagattgttgcagtagatcaggagggagc 272
          ||  : : :    |: ||:|:|   : : ::  : ||  |:  :: | :  :: ::| :
Db    777 TGTGKGKTKKKGTGKKGTKGKGKGGKKKKKKTGKKTTTGTKGGKKGGGKKTKKKKKGKKT 836

Qy    273 tcgtgttgttgggacgaactgtatgcttgctcgtggtggaactggtgctgtagcg 327
          | ::: |:|:::: :   |:: |  :|: : ::::::   ::| : ::  | |
Db    837 TKKKKGTKTKKKKTKKGGKTKKKTTGKTKKKKKKKKKKKGKGKKGGKTKKGGGKG 891


RESULT   3
CNS0039G/c
LOCUS       CNS0039G                 1101 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL063921
VERSION     AL063921.1  GI:4941778
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR08K10"
                     /note="end : TET3"
BASE COUNT      201 a     64 c    131 g    202 t    503 others
ORIGIN


  Query Match             1.5%;  Score 57.8;  DB 12;  Length 1101;
  Best Local Similarity   16.9%;  Pred. No. 0.0013;
  Matches  103;  Conservative  260;  Mismatches  245;  Indels    1;  Gaps    1;

Qy   1004 aactttctgaagtggtcctcaaaagaaatgaagatgaaaatggaaaaactgaagagactc 1063
          :: :|  :::: |: :  : ::|:::: |:|::: :::  | :: |  : ::: : | :
Db   1091 WDRDTRKDDWDWTKWWTWWKDRADDRRWAGDADRWAWDDGAGTWWTATWWWWWWWWATWD 1032

Qy   1064 ttgttgctgagcagtgcaatttgaccaaagatcctaacccaatgtctggaaaggaacgtg 1123
```

```
Db   1031 TWWDKWWWWATAAKTDTAWTWWRTAWRADWAGRDRGAGKRDRDAATDADGAGRRDGGRKR 972

Qy   1124 atcaggttgctgagcagtgcaatttgaccaaagatccgaaaccagtgtctgggcagaaat 1183

Db    971 KDKKDRKDGDDDKKGGKKKKAAKAAKWATKWWDDWDWDKDWKWDGAKDRKADDDDGAGDK 912

Qy   1184 gtgagcagatctgcaatgagccatgtgaagaagttgttctcaaaagaagctccaaatcta 1243

Db    911 DDDGKGKDADDDTDGTKDDDDKDKWDDWDKAKGTWGDATWAWAATDWWWWGWADADWWTW 852

Qy   1244 agaggaagacggataaga-agttgatgaagaagcagcagcacagcaagaaacgcactgcc 1302

Db    851 DAAADDWWADDRWDAWAWKWDDAWAWGARTADRRDWGDRAGKRGGARKRRDRKRADDKRD 792

Qy   1303 caggctgatgtttcagatgcaaagctttgtcggagaaagccaaaaaaggtgcggcttcta 1362

Db    791 AADDRDDAATWTTWTTTTTRDTDDWKWKTDTWTRWAADRTWDRDDDDDRDRAGTAGRKWR 732

Qy   1363 tcagaaattataaatgctaaccaggttgaggattctagaagtgacgaagttcatcgtgaa 1422

Db    731 RTWKRRWKRRDTRWDDADADDTARDDRRRRGDDGADAGKGKKTGRKRRRRDRATWDRTDA 672

Qy   1423 aatgccgctgatccctgtgaggatgatagaagtaccatcccggtcccgatggaagtaagc 1482

Db    671 WWADAAWWTTTDTDTDDWDKRDRRRKGARRRRRTTARAAWDWWTWKAWDWAKWDWKTRAD 612

Qy   1483 atggatattcctgttagcaaccatacagtgggagaagatgggttaaaatcaagtaagaac 1542

Db    611 RWDRWAADTWTDARKADRDWAKARAWRARRDRARAARADRRWTTKGKTTTATWTTWAARA 552

Qy   1543 aagacaaaacgcaaatactctgatgttgtagatgatggatcatcacttatgaactggctg 1602

Db    551 AWWAWWAWATTTATWTTTWTTTWTTTTTTTTTTTWTTTAAWWWAAWWTATWAAWTAAAWA 492

Qy   1603 aatggaaaa 1611

Db    491 AAAAAAWAA 483
```

```
RESULT   4
CNS0073W
LOCUS       CNS0073W                 922 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR14D09 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL066784
VERSION     AL066784.1  GI:4945247
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 922)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .922
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR14D09"
                     /note="end : TET3"
BASE COUNT      223 a     95 c     109 g     221 t     274 others
ORIGIN
```

```
  Query Match             1.4%;  Score 56.2;  DB 12;  Length 922;
  Best Local Similarity   21.6%;  Pred. No. 0.0031;
  Matches   66;  Conservative  120;  Mismatches  119;  Indels    1;  Gaps    1;

Qy     35 tgttctagcgtgtgcagcggctagctgattgattgtcttctgtgatatatccagagctcg 94

Db    617 KGTKKTTKKTKKKGTKKKKTKTTKGKKKGTGKGKKTGTTGGGKKKTKKTKKKKKTGKKKK 676

Qy     95 tgttttgtggtt-tgtggtttgtggtttgtgcttggattgttgatgtgctaattcgcggc 153

Db    677 KKTGKKGKKGTKKTKTKTTKGGKGGKKTKKKKTTKKTTKKTTKTKKKKKTTKTKKTGTKK 736

Qy    154 gttacaagatcactgctggattgatattgagttgtgcctcggctgtgctggctgtgtgtt 213

Db    737 KKTTKTTKKGKGTKKGKKGKKGGKTKGTGTKKKGTGGKGKKGKKGKKKKKKKKGKKKKTG 796

Qy    214 gattctctcctcgtcgtggtgatcgatatggagattgttgcagtagatcaggagggagct 273

Db    797 KGTGTKKKTKGTGKTKKTKKKKKKKKTKTKKKGTKKGGKTKKGTKKGKKKKGTKKKNKTT 856

Qy    274 cgtgttgttgggacgaactgtatgcttgctcgtggtggaactggtgctgtagcgccagtg 333

Db    857 KKTGTGTKKKKKKKKGKGKGKKKKGKKKGKKKGKGGKGKKKGKKGKKKGGGGGKGGGKK 916

Qy    334 ttggag 339

Db    917 GKKKKG 922
```

```
RESULT   5
CNS016LW/c
LOCUS       CNS016LW                 1101 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN16J16 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL106910
VERSION     AL106910.1  GI:5624430
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
```

```
                     /plasmid="pBeloBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN16J16"
                     /note="end : T7"
BASE COUNT       222 a     80 c    146 g    113 t    540 others
ORIGIN
```

Query Match 1.4%; Score 54; DB 12; Length 1101;
Best Local Similarity 13.1%; Pred. No. 0.012;
Matches 56; Conservative 113; Mismatches 257; Indels 0; Gaps 0;

```
Qy      60 tgattgattgtcttctgtgatatatccagagctcgtgttttgtggtttgtggtttgtggt 119
           ::| ::  :::|::|  ::| | |: ||::  :||:: :: : |: ::: :    |:
Db     750 KBANKKKCDDKCKDCAAWKAAAMAKACAKMCDKCGKKGKKNKGGKNKKKNKNNNNGKNNN 691

Qy     120 ttgtgcttggattgttgatgtgctaattcgcggcgttacaagatcactgctggattgata 179
           :: ::  : : ::|::        :       :     |  : :   :: ::      :
Db     690 KKNKKKGKNKGKKGKKNNGNGNKKNNGNGNNNKNNNGNCKNKGKKKKKKKKKNKNNKNNN 631

Qy     180 ttgagttgtgcctcggctgtgctggctgtgtgttgattctctcctcgtcgtggtgatcga 239
           :    |     : :| :::: :| ::::::::| :: : :     : :::   |   |
Db     630 NNKNNANGANANKKKGGKKKKNGKGKKKKKKKKKGKKKKKKKKKGKNKKKKNNAANKNA 571

Qy     240 tatggagattgttgcagtagatcaggagggagctcgtgttgttgggacgaactgtatgct 299
             : ::  ::       :  :      |:| ::: : :  :       :  :
Db     570 NNNANKKNKKNNKKKNNNNAKKNGKNNNNNNNCKCCKKKNKNKNNKNNNNNNNKCNKCNC 511

Qy     300 tgctcgtggtggaactggtgctgtagcgccagtgttggagctgacagcgacgcctcgtca 359
           ::   :   :   | : ::       :       | ::| : ::     :       |  : |
Db     510 KKACNKCCNKNNKAKKNKKNKNNAKKNNNNNNNGNKKGKKKKKKKKKKANNNNCNKKCCN 451

Qy     360 ggatgcagccgctgaagctggtgtagacgaaccggcacaacaccaatgcgagcatttctc 419
           : ::  : | | :     :::                 |   |  |     :   :
Db     450 NKKKKGKKKCCCNKKNCNCBKKNNNNNNCNKNNNNNKNGNCNNNANNGNCNNGNKGCNKG 391

Qy     420 cataagagggtatgttgctcttcttcagaagaaggatccaaaattctgctctctatctcg 479
             ||  ||| ||::: | || ||  |       |       ||   |     |||
Db     390 GNNNAGTNGGTGTGKKKKTTTTTTTTNGTTNTCCNNTTTNCTCCTCNTNTNNTNNTCTTT 331

Qy     480 gatttt 485
               ||||
Db     330 TTTTTT 325
```

```
RESULT   6
CNS006U0
LOCUS       CNS006U0                  884 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR14N21 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL065923
VERSION     AL065923.1  GI:4944891
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 884)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES              Location/Qualifiers
     source           1. .884
                      /organism="Drosophila melanogaster"
                      /db_xref="taxon:7227"
                      /clone_lib="RPCI-98"
                      /clone="BACR14N21"
                      /note="end : T7"
BASE COUNT       230 a     62 c    139 g    124 t    329 others
ORIGIN
```

Query Match 1.4%; Score 53.8; DB 12; Length 884;
Best Local Similarity 18.2%; Pred. No. 0.013;
Matches 46; Conservative 110; Mismatches 97; Indels 0; Gaps 0;

```
Qy      90 gctcgtgttttgtggtttgtggtttgtggtttgtgcttggattgttgatgtgctaattcg 149
           | : : |:|||:|: |:|:  :|:|:| |:|  || ::::   |  | :::: :  |  |
Db     596 GGKGKGGKTTTKTKTTKTKGTKTKTKTTGKTGTTGTKKKKGGGGGGGGKKKKGKGGTGKG 655

Qy     150 cggcgttacaagatcactgctggattgatattgagttgtgcctcggctgtgctggctgtg 209
             ||  |::    :          :: ||: : ::: ||:|:|  : ::  ||: ::| ||:|
Db     656 GGGGGKKKGGTKTGGKDGTGGKKKTTKKKKKKKKKGTKGKGKGKTKKKSGTKKKKGGTGKG 715

Qy     210 tgttgattctctcctcgtcgtggtgatcgatatggagattgttgcagtagatcaggaggg 269
           |||:|  | : :  : :  :::| : : | : ::| : :::::  :: :     : ::|
Db     716 TGTKGTGTKKKGGKGKGGKKGGKGKKGKKKKKGKKKKKKKKGKKKKKTGKKTKGKKG 775

Qy     270 agctcgtgttgttgggacgaactgtatgcttgctcgtggtggaactggtgctgtagcgcc 329
             : |::: ::|:::  :   ::: :: ::: : :::::::   :::|: ::: : :
Db     776 GTKKKGKKKGKKTKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKTKKKKKKKKKK 835

Qy     330 agtgttggagctg 342
           :::::::: : ::
Db     836 CKKKKKKKKKKK 848
```

```
RESULT   7
CNS00CNG/c
LOCUS       CNS00CNG                  939 bp    DNA     linear   GSS 04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR26H16 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL059400
VERSION     AL059400.1  GI:4946964
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 939)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
```

EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

```
FEATURES             Location/Qualifiers
     source          1. .939
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR26H16"
                     /note="end : TET3"
BASE COUNT       71 a    349 c    104 g    180 t    235 others
ORIGIN
```

```
  Query Match             1.4%;  Score 52.8;  DB 12;  Length 939;
  Best Local Similarity   14.3%;  Pred. No. 0.023;
  Matches   42;  Conservative  139;  Mismatches  113;  Indels    0;  Gaps    0;

Qy      44 gtgtgcagcggctagctgattgattgtcttctgtgatatatccagagctcgtgttttgtg 103
           ::: :  : :: : : :: ::: :::: : :::: : :       : : ::::: ::::
Db     565 KKKGKKKKKKDDKKKKKKKKKKKDKKKAKKKKKKKKKKKAKKKAKKKKKKKDDADKKK 506

Qy     104 gtttgtggtttgtggtttgtgcttggattgttgatgtgctaattcgcggcgttacaagat 163
           :::::::: :::::::::::::  ::: :::: : :::: :| :: : :: :::  | :|
Db     505 KKDKKKKKAKDDKKKKKKDKKDAKKKKKKKKKAKKKKKKKKAKKKKKKKKKKKKKAKKAA 446

Qy     164 cactgctggattgatattgagttgtgcctcggctgtgctggctgtgtgttgattctctcc 223
           :: ::   :| :|:|: :  :::  :  : :::: ::: : ::::::: :: : :
Db     445 KKKKKKKKAKAKGKKAKTDKKAAKKKKDKKAKKKKKKKKKKDKAKKKKKKKKKKKKKKK 386

Qy     224 tcgtcgtggtgatcgatatggagattgttgcagtagatcaggagggagctcgtgttgttg 283
           : |::: :   |   :::| : ::::|  | | :   || |||||   | :: | ||
Db     385 GGAKDGKKKGKGGGGGGGKKRAADRKKKKGAGGGATRGGKGGGGGGAGAGGGGKKGGATG 326

Qy     284 ggacgaactgtatgcttgctcgtggtggaactggtgctgtagcgccagtgttgg 337
           ||: |||  |  :|  ::   | || |||: :|: |  |    |  : ::::||
Db     325 GGRAGAAGGGGKKGGGKRRRTGGGGGGGARGKGRGGGAGGGAAGGKWTDKKKGG 272
```

```
RESULT   8
CNS00DG1/c
LOCUS       CNS00DG1                1101 bp    DNA     linear   GSS 04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR27J03 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL069971
VERSION     AL069971.1  GI:4947993
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR27J03"
                     /note="end : TET3"
BASE COUNT      227 a    174 c     33 g    361 t    306 others
ORIGIN
```

```
  Query Match             1.3%;  Score 51.8;  DB 12;  Length 1101;
  Best Local Similarity   30.2%;  Pred. No. 0.045;
  Matches  149;  Conservative   96;  Mismatches  248;  Indels    0;  Gaps    0;

Qy    1127 aggttgctgagcagtgcaatttgaccaaagatccgaaaccagtgtctgggcagaaatgtg 1186
           | ::|| : :: :|:| | |: |    || |   :||| |: :   | :  |||| |||
Db     972 AARKTGRWAWRRRGGGGATTKAGGWAGAAAARGARAAAAAARARRAGGAKTGGAAAGGTG 913

Qy    1187 agcagatctgcaatgagccatgtgaagaagttgttctcaaaagaagctccaaatctaaga 1246
             :|:  |      ||  : |  :| |::|| :|:: ||||:: ::|: ||| : || :
Db     912 GAARGRGATTGKTAAAGRGRGGATRAAAWVTTTWTSWAAAAAVDTSSTSAAAAASAAAAV 853

Qy    1247 ggaagacggataagaagttgatgaagaagcagcagcacagcaagaaacgcactgcccagg 1306
             :|| |   | || ||    :  || ||  |  |     : || |||:::|:  :  :
Db     852 ARAAAAAAAAAAAAAAAAAAAWAAAAAAATWAAAAAAASTSKTAAAAAASSSAVAASAWWWA 793

Qy    1307 ctgatgtttcagatgcaaagctttgtcggagaaagccaaaaaaggtgcggcttctatcag 1366
             :  |   ::| : : :    | |   | ||||  :||||| : : : ::   | :
Db     792 TGSSRATAAAWRARSARGRAAARTWTAAAAAAAAGAARAAAAAASSVASASWAWRAVSSA 733

Qy    1367 aaattataaatgctaaccaggttgaggattctagaagtgacgaagttcatcgtgaaaatg 1426
             || :  |:|    ||    | | |:| |   |||| :   :| || |:  :::||  :
Db     732 SAAAWSAAWAAARVAAAATTGATAARGGTGGGAGAAAARTTWWAATTAAWGWWRRAATGR 673

Qy    1427 ccgctgatccctgtgaggatgatagaagtaccatcccggtcccgatggaagtaagcatgg 1486
               :|       |::|  :|: |   |  ||          :|   | ||:
Db     672 TAWRARARAATATGAARSAGARTWAATWRAAAATGWAAAATTTARTTTGAATARATTSAA 613

Qy    1487 atattcctgttagcaaccatacagtgggagaagatgggttaaaatcaagtaagaacaaga 1546
           |          | :: : |  :|:  : |    :  : |  |||  |   || :| |  |
Db     612 ASSSAWTAASAAAVWTSAAASSASAASTAASSAWAASATSSAAASAATTAAAAWAAATAA 553

Qy    1547 caaaacgcaaatactctgatgttgtagatgatggatcatcacttatgaactggctgaatg 1606
             |||::: ||| ::    | ::|  | |  |  :: ::|:| |    ||:: ::|: |
Db     552 AAAAWVSAAAAAVSSASAAAVWTAAAAARAAAASVAVVTSATTSSAAAASWTSSTSSASA 493

Qy    1607 gaaaaaagaaaag 1619
           :   || :||:|:
Db     492 STSSAATVAAVAS 480
```

```
RESULT   9
CNS005TE/c
LOCUS       CNS005TE                997 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR12K22 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL060767
VERSION     AL060767.1  GI:4943573
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 997)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
```

```
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .997
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR12K22"
                     /note="end : TET3"
BASE COUNT       89 a     99 c     13 g    258 t    538 others
ORIGIN
```

```
  Query Match             1.3%;  Score 50.4;  DB 12;  Length 997;
  Best Local Similarity   18.3%;  Pred. No. 0.1;
  Matches   88;  Conservative  165;  Mismatches  227;  Indels    0;  Gaps    0;

Qy   1149 gaccaaagatccgaaaccagtgtctgggcagaaatgtgagcagatctgcaatgagccatg 1208
          :|  |:|||   :| | || |    :  :::|: :   : ||:   |  | | :    :
Db    957 RARGARAGARAARAGARRAGAGARARAAAARRRARRRRAGRRAGRAGRGRSAGGGRGRGGR 898

Qy   1209 tgaagaagttgttctcaaaagaagctccaaatctaagaggaagacggataagaagttgat 1268
            || :|  :      |:|: :::    ||:   |::::|:::: |:| :::|:|  |:
Db    897 RAAAARAARDARARARARARARRRARAKAARARGARRRRGRRRRRGRAGRRRARGAGGRA 838

Qy   1269 gaagaagcagcagcacagcaagaaacgcactgcccaggctgatgtttcagatgcaaagct 1328
          |::|: : :: |: : :  ||:|:| | :      : :  :| :    |:  : ::::
Db    837 GRRGRGRGRRRARARRRAGAARARARGRRARAAARARARRRRAGRRRRRARGARGRRRRGG 778

Qy   1329 ttgtcggagaaagccaaaaaaggtgcggcttctatcagaaattataaatgctaaccaggt 1388
            :  ||::: ::  ||::|: : : |      :  |::|   | :|: :  ::  :|
Db    777 RRRGRGGRRRGRRRRAARRARARARAGARAAARRRRARRAGAAARRARRRRRRRRRGAG 718

Qy   1389 tgaggattctagaagtgacgaagttcatcgtgaaaatgccgctgatccctgtgaggatga 1448
            :|::|    ::||: ||  |::   :  :   |:: |  :  ::     : ::|:: :
Db    717 ARARRARRRARRAARAGAGAARRRRGRRGRGAGARRRGRGRGRRRGAGARRRRRGRRMRT 658

Qy   1449 tagaagtaccatcccggtcccgatggaagtaagcatggatattcctgttagcaaccatac 1508
            : ::: :  :     :    || ::::| : : : |:|       |  ::  :  : |
Db    657 RRARRRRRARRGAGAARRRARGARRRRRGRRGRRRRGRARGRRRRAGRGRRRGRRARGAR 598

Qy   1509 agtgggagaagatgggttaaaatcaagtaagaacaagacaaaacgcaaatactctgatgt 1568
          :| ::  :::::  :    |::  ::: :|::| :  | :::| : ::: :    ::
Db    597 RGRRRAGRRRRRRARAGAGARRGRRRRGRARRAARGAARRRRARRARRRRRRAAARRGAG 538

Qy   1569 tgtagatgatggatcatcacttatgaactggctgaatggaaaaaagaaaagaactggaag 1628
            |:| |: |:|  |  |   : :::   :  | : ||:::||  :::: |:   :|::
Db    537 RAGARARGRAGRARRARRARRRRRRRGGARRRGGRRGGRRRAAGARRRRAARAAARARR 478
```

```
RESULT  10
CNS00JTL
LOCUS       CNS00JTL                 909 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC:
            BACR39G08 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL076720
```

```
VERSION     AL076720.1  GI:4956298
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 909)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .909
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR39G08"
                     /note="end : TET3"
BASE COUNT       293 a     55 c    175 g    144 t    242 others
ORIGIN
```

```
  Query Match             1.2%;  Score 48.6;  DB 12;  Length 909;
  Best Local Similarity   24.4%;  Pred. No. 0.28;
  Matches   79;  Conservative  106;  Mismatches  138;  Indels    1;  Gaps    1;

Qy     15 caaacgcaaaccccccaggttgttctagcgtgtgcagcggctagctgattgattgtcttc 74
          ::: : ::::::::::   || || |   :|||     :  |   :: ||  |::| |:
Db    464 MMMCMCMMMHYYYYYYTTTTTTTTTTTTKTGTTTTTTKTWTTTTKKTTTTTTKKTKTKT 523

Qy     75 tgtgatatatccagagctcgtgttttgtggtttgtggtttgtggtttgtgcttggattgt 134
          | :: | | :     : : :| ||||:::|: :| :| ||: :|:::|||  :|: :||
Db    524 TTBKTTTTTKTGTTTKTKKKTTTTTTKKKGKGKGGKGGTTKGKGKKKGTGKGKGKTKTGG 583

Qy    135 tgatgtgctaattcgcggcgttacaagatcactgctggattgatattgagttgtgcctcg 194
          ::  |:| |   : : :|  :|    | :   :|  :: | : |  |: :||||   | :
Db    584 KKGGGKGKTKKGKTKGKGKTKTKKTGGGKKKGKGTGKKGTGKTTGGTKGKTTGTTKKTTK 643

Qy    195 -gctgtgctggctgtgtgttgattctctcctcgtcgtggtgatcgatatggagattgttg 253
           : |||| :|| :::| :: | :: :       |  |:: :|   : : :|: :   : :|
Db    644 TKTTGTGGKGGGKKKGGKKGGKKKTKGGGKGKGGTGKKTKGGGGKGKGKGKKKKGGKGKG 703

Qy    254 cagtagatcaggagggagctcgtgttgttgggacgaactgtatgcttgctcgtggtggaa 313
            :| : :  :: |:| : :  |:||:||:||  :    ||  : ::  | :: :|::
Db    704 TTKTTKKKKTKKTGKGKKTKTTTKTTKTTKGGKGKGGGGGTKGKTKKTGTTKKTKTKKKK 763

Qy    314 ctggtgctgtagcgccagtgttgg 337
            :::  |:|     : :: |:::
Db    764 KNKKKTKTKTGTKTGMGKKTTKKK 787
```

```
RESULT  11
CNS017TG
LOCUS       CNS017TG                 985 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN37G05 of DrosBAC library from Drosophila melanogaster (fruit
```

```
          fly), genomic survey sequence.
ACCESSION   AL108478
VERSION     AL108478.1  GI:5628782
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 985)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .985
                     /organism="Drosophila melanogaster"
                     /plasmid="pBeloBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN37G05"
                     /note="end : SP6"
BASE COUNT      139 a    188 c    248 g    115 t    295 others
ORIGIN


  Query Match             1.2%;  Score 47.4;  DB 12;  Length 985;
  Best Local Similarity   36.1%;  Pred. No. 0.59;
  Matches   57;  Conservative   42;  Mismatches   59;  Indels    0;  Gaps    0;

Qy      32 ggttgttctagcgtgtgcagcggctagctgattgattgtcttctgtgatatatccagagc 91
           | ||||: |:: :||:|  | ||   | |  ||  ||:: || ::|| | : |   | |
Db     811 GTTTGTWTTMKTKTGKGGGGGGGKGKGTTWTTTTTTTRKTTTTKRTGTTTKTTTTTGKGT 870

Qy      92 tcgtgttttgtggtttgtggtttgtggtttgtgcttggattgttgatgtgctaattcgcg 151
           | |||:||: | || :|::::| :: : :|::: ::|  :|||  ::|: :  :||: |
Db     871 TTGTGKTTKTTTGTAKGKKKKTGKKTKGKTKKKGKKGTGKTGTTTGKKTKTKTGKTCKTG 930

Qy     152 gcgttacaagatcactgctggattgatattgagttgtg 189
             :::    | |   :: ||  ||  : |:   ||::
Db     931 TGKKKGTGKGKTKTKKKGTGTKTTTKKTTKNTTTGKK 968


RESULT  12
CNS00LO0/c
LOCUS       CNS00LO0                1101 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC:
            BACR32D23 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL068607
VERSION     AL068607.1  GI:4958689
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR32D23"
                     /note="end : TET3"
BASE COUNT      155 a    166 c      7 g    284 t    489 others
ORIGIN


  Query Match             1.2%;  Score 46.4;  DB 12;  Length 1101;
  Best Local Similarity   21.0%;  Pred. No. 1.1;
  Matches   65;  Conservative  112;  Mismatches  132;  Indels    0;  Gaps    0;

Qy      75 tgtgatatatccagagctcgtgttttgtggtttgtggtttgtggtttgtgcttggattgt 134
           |:::  :: :   |::    ::|||: ||:| | :: :| ::::: :::| : :::|:||
Db     810 TKKKKGRKKKKGTGDKKRTWWKTTTKTTGKTGTAWKTKTGKKKKKGKKKGKKGDWTKGT 751

Qy     135 tgatgtgctaattcgcggcgttacaagatcactgctggattgatattgagttgtgcctcg 194
           : :|:|: :  :| : || ::| :    : :    : : |:: | ::| | :  :  : :
Db     750 KWDTKTKTKGKKTGSGGGKKKTKYKGTKKTWWAAWKAKKTKKTTKKKGTGAKTADKTKKK 691

Qy     195 gctgtgctggctgtgtgttgattctctcctcgtcgtggtgatcgatatggagattgttgc 254
           : :::| | | ::::||:  ::    |  : :: :     :  :  |:  :|:|: :
Db     690 KKKDKGKTTGKWKKKTGKAAWWGTGDTRTKDKKKKAWAAWNDTTDGKTKAKKADTKADTG 631

Qy     255 agtagatcaggagggagctcgtgttgttgggacgaactgtatgcttgctcgtggtggaac 314
           : |:|| | :|:| :    ||:|:: : |   :|: |:  :|   |   ::|:|::
Db     630 KKGAKATKAAKAKGWDADADGTDTKKAKTGTKTKADDTKGGKGAARGKGGRKGRTKDKGA 571

Qy     315 tggtgctgtagcgccagtgttggagctgacagcgacgcctcgtcaggatgcagccgctga 374
           :|::| |  |: |  |::|: :| :  :: :|  : :  : ::   :     :  | |:
Db     570 KGDKGATAGAKKGKTAKKGDAKGKKDGKRADGGARTKRGKKKKAGTKKGTKKDKKGGTDK 511

Qy     375 agctggtgt 383
           | |:|||:
Db     510 KGATRGTGK 502


RESULT  13
CNS017WI/c
LOCUS       CNS017WI                1101 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN37L10 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL108588
VERSION     AL108588.1  GI:5628892
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
```

```
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /plasmid="pBeloBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN37L10"
                     /note="end : T7"
BASE COUNT      293 a    177 c    105 g    186 t    340 others
ORIGIN
```

```
  Query Match             1.2%;  Score 46.2;  DB 12;  Length 1101;
  Best Local Similarity   33.9%;  Pred. No. 1.3;
  Matches   84;  Conservative   39;  Mismatches  125;  Indels    0;  Gaps    0;

Qy      63 ttgattgtcttctgtgatatatccagagctcgtgttttgtggtttgtggtttgtggtttg 122
           ||  || | || | |  | | |  |       ||||:: |  ||| |: ||| |  ||||
Db     510 TTTTTTTTTTNTTTTTTTTTTKKATNAANGNTGTTKKTTTTTTTTTKTTTTTTTWTTTG 451

Qy     123 tgcttggattgttgatgtgctaattcgcggcgttacaagatcactgctggattgatattg 182
           |  ||   |: ||  | |:       | |  :::   | :     |  |   |  |  |
Db     450 TTTTTTTTTKNTTTTTTKGGGGGGGGGGTTKKKGGGAARNNNNGGGGGTNNTTTTTNTT 391

Qy     183 agttgtgcctcggctgtgctggctgtgtgttgattctctcctcgtcgtggtgatcgatat 242
           |||     |    |:|  |      | ::::: :: |       | ||:||    | | |
Db     390 GGTTTNTTTTTTTTTKTTKTNTGGTTTKKKKKKKKGTNNTGNNNTGGTKGTTTGNGNTNT 331

Qy     243 ggagattgttgcagtagatcaggagggagctcgtgttgttgggacgaactgtatgcttgc 302
           |: | || :||   : : :  :: ::| : : |:: ::|| |:  |   |:|  | |::
Db     330 GKNGKTTTKTGTTTKKKKKKKKKGKKGKKGKGGKKGKKTTTGKKKGKKKTKTTGGGTKKK 271

Qy     303 tcgtggtg 310
           | ::||||
Db     270 TTKKGGTG 263
```

```
RESULT  14
AW684844
LOCUS       AW684844                706 bp    mRNA    linear   EST 15-JUN-2000
DEFINITION  NF022B10NR1F1000 Nodulated root Medicago truncatula cDNA clone
            NF022B10NR 5', mRNA sequence.
ACCESSION   AW684844
VERSION     AW684844.1  GI:7559580
KEYWORDS    EST.
SOURCE      barrel medic.
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 706)
  AUTHORS   Watson,B.S., Shin,H.-S., Lopez-Meyer,M., Scott,A.D., Harris,A.R.,
            Gonzales,R.A., Bell,C.J., Flores,H.R., Inman,J.T., Weller,J.W., May
            ,G.D. and Paiva,N.L.
  TITLE     Expressed Sequence Tags from the Samuel Roberts Noble Foundation
            Medicago truncatula nodulated root library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Paiva NL
            Plant Biology Division
            The Samuel Roberts Noble Foundation
            2510 Sam Noble Parkway, Ardmore, OK 73402, USA
            Tel: 580 221 7317
            Fax: 580 221 7380
            Email: nlpaiva@noble.org
            Insert Length: 706   Std Error: 0.00
            Plate: 022  row: B  column: 10
            Seq primer: TCACACAGGAAACAGCTATGAC.
FEATURES             Location/Qualifiers
     source          1. .706
                     /organism="Medicago truncatula"
                     /db_xref="taxon:3880"
                     /clone="NF022B10NR"
                     /clone_lib="Nodulated root"
                     /tissue_type="root"
                     /dev_stage="Pooled developmental"
                     /note="Vector: Lambda Zap; Four-week old Rhizobium
                     meliloti-inoculated Medicago truncatula roots, containing
                     a mixture of young and old roots and nodules."
BASE COUNT      240 a    119 c    175 g    168 t      4 others
ORIGIN
```

```
  Query Match             1.2%;  Score 46;  DB 9;  Length 706;
  Best Local Similarity   50.9%;  Pred. No. 1.2;
  Matches  109;  Conservative    0;  Mismatches  105;  Indels    0;  Gaps    0;

Qy     997 actccaaaactttctgaagtggtcctcaaaagaaatgaagatgaaaatggaaaaactgaa 1056
           |||  |||   | ||||    |   |||      ||||||||||  |    || |  |||
Db     140 ACTGAAAAGACTACTGAGAAAGAGATCAGTGATGATGAAGATGATGAGCCCAAGAAGGAA 199

Qy    1057 gagactcttgttgctgagcagtgcaatttgaccaaagatcctaacccaatgtctggaaag 1116
           ||      ||| |  ||| |     ||  |  ||| ||     |  | | |     ||||
Db     200 GAAGAAGGTGTGGTCGAGGAAGTTGATGAGGACAAGGAGAAAGATGCCAAGAAGAAAAAG 259

Qy    1117 gaacgtgatcaggttgctgagcagtgcaatttgaccaaagatccgaaaccagtgtctggg 1176
           |   | |  ||||| || |  ||||  |  | | |||  | | ||| ||| | |    |
Db     260 AAGATTAAGGAGGTTTCTCATGAGTGGGAACTCATCAACAAGCAGAAGCCAATCTGGCTG 319

Qy    1177 cagaaatgtgagcagatctgcaatgagccatgtg 1210
           |  ||    ||  |||||  ||| ||   || ||
Db     320 CGCAAGCCAGAAGAGATCACCAAGGATGAATATG 353
```

```
RESULT  15
BI176637/c
LOCUS       BI176637                748 bp    mRNA    linear   EST 09-JUL-2001
DEFINITION  EST517582 cSTE Solanum tuberosum cDNA clone cSTE5K22 5' sequence,
            mRNA sequence.
ACCESSION   BI176637
VERSION     BI176637.1  GI:14642448
KEYWORDS    EST.
SOURCE      potato.
  ORGANISM  Solanum tuberosum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE   1  (bases 1 to 748)
  AUTHORS   van der Hoeven,R., Bezzerides,J., Bachem,C., Visser,R., Cho,J.,
            Chiemingo,A., Bougri,O., Buell,C.R., Ronning,C., Tanksley,S. and
            Baker,B.
  TITLE     Generation of ESTs from in vitro grown microtubers
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Cathy Ronning
            The Institute for Genomic Research
            For clone info: please contact Research Genetics, Libraries
            Division tel 1-800-711-6195, email cdna@resgen.com
            Seq primer: M13F-R.
FEATURES             Location/Qualifiers
     source          1. .748
                     /organism="Solanum tuberosum"
                     /cultivar="Bintje"
                     /db_xref="taxon:4113"
                     /clone="cSTE5K22"
                     /clone_lib="cSTE"
                     /tissue_type="axillary buds of stem explants; growing
                     sink-tubers"
```

/dev_stage="7, 8 and 10 days"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; Tissue supplied by Christian Bachem and Richard Visser (Department of Plant Breeding, Wageningen University, The Netherlands). The cSTA libraries will attempt to capture the induction and initiation/initial growth of the tuber in an in vitro system as described in Bachem et al. (Plant Journal, 1996). Small microtubers develop from axillary buds attached to stem explants when placed on a high sucrose medium (10%). Visible morphological changes occur synchronously at day five in the axillary buds. The first library, cSTA (1-20) consists of axillary buds harvested on days 1-3. This targets those genes involved in induction of the microtubers. The following libraries, cSTA (21-40) and cSTA (41-60), capture genes involved in tuber initiation and outgrowth. This library is noted as P3 in Tanksley lab notebooks."

BASE COUNT 208 a 236 c 131 g 173 t
ORIGIN

Query Match 1.2%; Score 45.6; DB 10; Length 748;
Best Local Similarity 52.0%; Pred. No. 1.6;
Matches 102; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

```
Qy    207 gtgtgttgattctctcctcgtcgtggtgatcgatatggagattgttgcagtagatcagga 266
          | |||||| | ||      || |  |   | | | | |   ||||||| |  | |   |
Db    457 GGGTGTTGTTGCTGGTGCTGTTGCTGGTGTTGTTGTTGGTGTTGTTGCTGGGGTTGTTGC 398

Qy    267 gggagctcgtgttgttgggacgaactgtatgcttgctcgtggtggaactggtgctgtagc 326
          || | |  || || ||          |  |  ||||| | | ||   |||||| ||| ||
Db    397 TGGTGTTGTTGCTGGTGTTGTTGCTGGGGTTGTTGCTGGGGTTGTTGCTGGTGTTGTTGC 338

Qy    327 gccagtgttggagctgacagcgacgcctcgtcaggatgcagccgctgaagctggtgtaga 386
            | |  | |   || ||||| ||  |  |   | ||  || ||||  |||| ||  |
Db    337 TGGAATTGTTGCTGTGGCAGCGGCGGGTGCTGGTGCTGGTGCTGCTGGTGCTGCTGGTGC 278

Qy    387 cgaaccggcacaacac 402
          ||   ||  || |||
Db    277 TGATGAGGATCACCAC 262
```

Search completed: September 16, 2002, 18:12:16
Job time: 9974 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: September 16, 2002, 22:23:21 ; Search time 88.88 Seconds
(without alignments)
2057.332 Million cell updates/sec

Title: US-09-828-068-2
Perfect score: 5526
Sequence: 1 MEIVAVDQEGARVVGTNCML.....NKNPADFTTISNDNEYMDYR 1057

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	5526	100.0	1385	10	Q9LWL4	Q9lwl4 oryza sativ
2	5516	99.8	1057	10	Q947D1	Q947d1 oryza sativ
3	393	7.1	1096	10	Q9LYD9	Q9lyd9 arabidopsis
4	221	4.0	1983	4	Q9Y4D6	Q9y4d6 homo sapien
5	216	3.9	2081	10	Q9LH98	Q9lh98 arabidopsis
6	215.5	3.9	3507	5	Q23587	Q23587 caenorhabdi
7	215	3.9	1690	5	O44929	O44929 drosophila
8	209.5	3.8	852	10	Q9SZ55	Q9sz55 arabidopsis
9	208.5	3.8	1295	5	Q22257	Q22257 caenorhabdi
10	207.5	3.8	1520	12	Q9QJ16	Q9qj16 human herpe
11	206	3.7	3484	5	P91257	P91257 caenorhabdi
12	205.5	3.7	2083	5	Q9N435	Q9n435 caenorhabdi
13	204.5	3.7	5327	5	O76891	O76891 drosophila
14	204	3.7	2768	5	Q9VC00	Q9vc00 drosophila
15	204	3.7	2951	5	Q9W3Z0	Q9w3z0 drosophila
16	203	3.7	1047	5	Q9GRK7	Q9grk7 leishmania
17	202.5	3.7	2271	16	Q99QY4	Q99qy4 staphylococ
18	202	3.7	2062	4	Q9H231	Q9h231 homo sapien
19	201.5	3.6	665	5	O96229	O96229 plasmodium
20	201.5	3.6	1784	5	Q9VE02	Q9ve02 drosophila
21	200.5	3.6	3254	5	Q9BK45	Q9bk45 plasmodium
22	200	3.6	2647	5	Q9U4X0	Q9u4x0 plasmodium
23	200	3.6	3111	5	Q9VH10	Q9vh10 drosophila
24	199.5	3.6	1302	2	Q49547	Q49547 mycoplasma
25	199.5	3.6	17352	5	Q95YM2	Q95ym2 procambarus
26	199	3.6	1230	5	Q20626	Q20626 caenorhabdi
27	198.5	3.6	1043	10	O82345	O82345 arabidopsis
28	198.5	3.6	1690	5	Q9VJE5	Q9vje5 drosophila
29	198	3.6	615	6	O77733	O77733 saguinus oe
30	197.5	3.6	1852	3	Q9C2H4	Q9c2h4 neurospora
31	196.5	3.6	963	11	Q99MX4	Q99mx4 glaucomys v
32	196.5	3.6	1579	13	Q90X96	Q90x96 xenopus lae
33	196	3.5	1513	12	Q9WSZ8	Q9wsz8 human herpe
34	196	3.5	2910	10	Q9FND5	Q9fnd5 arabidopsis
35	195	3.5	667	4	Q96D08	Q96d08 homo sapien
36	195	3.5	1616	4	Q96PH3	Q96ph3 homo sapien
37	195	3.5	1730	5	Q9VC59	Q9vc59 drosophila
38	193.5	3.5	1803	5	Q9VU16	Q9vu16 drosophila
39	192	3.5	3201	5	Q9W0U2	Q9w0u2 drosophila
40	192	3.5	6632	5	Q17362	Q17362 caenorhabdi
41	192	3.5	6642	5	O01761	O01761 caenorhabdi
42	191.5	3.5	990	13	Q91803	Q91803 xenopus lae
43	191.5	3.5	2158	4	Q9Y6Y3	Q9y6y3 homo sapien
44	191.5	3.5	2265	4	Q9Y6Y4	Q9y6y4 homo sapien
45	191.5	3.5	3394	4	Q9Y6V0	Q9y6v0 homo sapien

ALIGNMENTS

RESULT 1
Q9LWL4
ID Q9LWL4 PRELIMINARY; PRT; 1385 AA.
AC Q9LWL4;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL PROTEIN.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT clone:P0485D09.";
RL Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AP001859; BAA94774.1; -.
KW Hypothetical protein.
SQ SEQUENCE 1385 AA; 151678 MW; 16E784264EEFC7B5 CRC64;

Query Match 100.0%; Score 5526; DB 10; Length 1385;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1057; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MEIVAVDQEGARVVGTNCMLARGGTGAVAPVLELTATPRQDAAAEAGVDEPAQHQCEHFS 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     329 MEIVAVDQEGARVVGTNCMLARGGTGAVAPVLELTATPRQDAAAEAGVDEPAQHQCEHFS 388

Qy      61 IRGYVALLQKKDPKFCSLSRIFHDQKKCDEHKASSSPFSVAKFRRWDCSKCLDKLKTSDN 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     389 IRGYVALLQKKDPKFCSLSRIFHDQKKCDEHKASSSPFSVAKFRRWDCSKCLDKLKTSDN 448

Qy     121 GTAPRTLPAKQNGTSDGCSITFVRSTFVPASVGSQKVSPSTQSSQGKNADRSTLPKSVQE 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    449 GTAPRTLPAKQNGTSDGCSITFVRSTFVPASVGSQKVSPSTQSSQGKNADRSTLPKSVQE 508

Qy    181 GNDSKCNAPSGKNGAAEANTDSPMKDLQGPAQNYDVAANVSEDNTSVDVGALPEVPQITW 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    509 GNDSKCNAPSGKNGAAEANTDSPMKDLQGPAQNYDVAANVSEDNTSVDVGALPEVPQITW 568

Qy    241 HIEVNGADQPPSTPKLSEVVLKRNEDENGKTEETLVAEQCNLTKDPNPMSGKERDQVAEQ 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    569 HIEVNGADQPPSTPKLSEVVLKRNEDENGKTEETLVAEQCNLTKDPNPMSGKERDQVAEQ 628

Qy    301 CNLTKDPKPVSGQKCEQICNEPCEEVVLKRSSKSKRKTDKKLMKKQQHSKKRTAQADVSD 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    629 CNLTKDPKPVSGQKCEQICNEPCEEVVLKRSSKSKRKTDKKLMKKQQHSKKRTAQADVSD 688

Qy    361 AKLCRRKPKKVRLLSEIINANQVEDSRSDEVHRENAADPCEDDRSTIPVPMEVSMDIPVS 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    689 AKLCRRKPKKVRLLSEIINANQVEDSRSDEVHRENAADPCEDDRSTIPVPMEVSMDIPVS 748

Qy    421 NHTVGEDGLKSSKNKTKRKYSDVVDDGSSLMNWLNGKKKRTGSVHHTVAHPAGNLSNKKV 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    749 NHTVGEDGLKSSKNKTKRKYSDVVDDGSSLMNWLNGKKKRTGSVHHTVAHPAGNLSNKKV 808

Qy    481 TPTASTQHDDENDTENGLDTNMHKTDVCQHVSEISTQRCSSKGKTAGLSKGKTHSAASTK 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    809 TPTASTQHDDENDTENGLDTNMHKTDVCQHVSEISTQRCSSKGKTAGLSKGKTHSAASTK 868

Qy    541 YGGESTRNGQNIHVLSAEDQCQMETENSVLSHSAKVSPAEHDIQIMSDLHEQSLPKKKKK 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    869 YGGESTRNGQNIHVLSAEDQCQMETENSVLSHSAKVSPAEHDIQIMSDLHEQSLPKKKKK 928

Qy    601 QKLEVTREKQTMIDDIPMDIVELLAKNQHERQLMTETDCSDINRIQSKTTADDDCVIVAA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    929 QKLEVTREKQTMIDDIPMDIVELLAKNQHERQLMTETDCSDINRIQSKTTADDDCVIVAA 988

Qy    661 KDGSDYASSVFDTNSQQKSLASQSTQKELQGHLALTTQESPHPQNFQSTQEQQTHLRMEE 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    989 KDGSDYASSVFDTNSQQKSLASQSTQKELQGHLALTTQESPHPQNFQSTQEQQTHLRMEE 1048

Qy    721 MVTIAASSPLFSHHDDQYIAEAPTEHWGRKDAKKLTWEQFKATTRNSPAATCGAQFRPGI 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1049 MVTIAASSPLFSHHDDQYIAEAPTEHWGRKDAKKLTWEQFKATTRNSPAATCGAQFRPGI 1108

Qy    781 QAVDLTSTHVMGSSSNYASRQPVIAPLDRYAERAVNQVHARNFPSTIATMEASKLCDRRN 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1109 QAVDLTSTHVMGSSSNYASRQPVIAPLDRYAERAVNQVHARNFPSTIATMEASKLCDRRN 1168

Qy    841 AGQVVLYPKESMPATHLLRMMDPSTLASFPNYGTSSRNQMESQLHNSQYAHNQYKGSTST 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1169 AGQVVLYPKESMPATHLLRMMDPSTLASFPNYGTSSRNQMESQLHNSQYAHNQYKGSTST 1228

Qy    901 SYGSNLNGKIPLTFEDLSRHQLHDLHRPLRPHPRVGVLGSLLQKEIANWSENCGTQSGYK 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1229 SYGSNLNGKIPLTFEDLSRHQLHDLHRPLRPHPRVGVLGSLLQKEIANWSENCGTQSGYK 1288

Qy    961 LGVSTGITSHQMNRKEHFEALNSGMFSAKWNALQLGSVSSSADFLSARNSIAQSWTRGKG 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1289 LGVSTGITSHQMNRKEHFEALNSGMFSAKWNALQLGSVSSSADFLSARNSIAQSWTRGKG 1348

Qy   1021 KMVHPLDRFVRQDICITNKNPADFTTISNDNEYMDYR 1057
          |||||||||||||||||||||||||||||||||||||
Db   1349 KMVHPLDRFVRQDICITNKNPADFTTISNDNEYMDYR 1385


RESULT   2
Q947D1
ID   Q947D1      PRELIMINARY;      PRT;  1057 AA.
AC   Q947D1;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   EMBRYONIC FLOWER 1-LIKE PROTEIN.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21380417; PubMed=11487698;
RA   Aubert D., Chen L., Moon Y.H., Martin D., Castle L.A., Yang C.H.,
RA   Sung Z.R.;
RT   "Emf1, a novel protein involved in the control of shoot architecture
RT   and flowering in arabidopsis.";
RL   Plant Cell 13:1865-1875(2001).
DR   EMBL; AF326768; AAK98529.1; -.
SQ   SEQUENCE   1057 AA;  116447 MW;  2A44FA15C93C3DA0 CRC64;


  Query Match           99.8%;  Score 5516;  DB 10;  Length 1057;
  Best Local Similarity 99.8%;  Pred. No. 0;
  Matches 1055;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 MEIVAVDQEGARVVGTNCMLARGGTGAVAPVLELTATPRQDAAAEAGVDEPAQHQCEHFS 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MEIVAVDQEGARVVGTNCMLARGGTGAVAPVLELTATPRQDAAAEAGVDEPAQHQCEHFS 60

Qy     61 IRGYVALLQKKDPKFCSLSRIFHDQKKCDEHKASSSPFSVAKFRRWDCSKCLDKLKTSDN 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 IRGYVALLQKKDPKFCSLSRIFHDQKKCDEHKASSSPFSVAKFRRWDCSKCLDKLKTSDN 120

Qy    121 GTAPRTLPAKQNGTSDGCSITFVRSTFVPASVGSQKVSPSTQSSQGKNADRSTLPKSVQE 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GTAPRTLPAKQNGTSDGCSITFVRSTFVPASVGSQKVSPSTQSSQGKNADRSTLPKSVQE 180

Qy    181 GNDSKCNAPSGKNGAAEANTDSPMKDLQGPAQNYDVAANVSEDNTSVDVGALPEVPQITW 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GNDSKCNAPSGKNGAAEANTDSPMKDLQGPAQNYDVAANVSEDNTSVDVGALPEVPQITW 240

Qy    241 HIEVNGADQPPSTPKLSEVVLKRNEDENGKTEETLVAEQCNLTKDPNPMSGKERDQVAEQ 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 HIEVNGADQPPSTPKLSEVVLKRNEDENGKTEETLVAEQCNLTKDPNPMSGKERDQVAEQ 300

Qy    301 CNLTKDPKPVSGQKCEQICNEPCEEVVLKRSSKSKRKTDKKLMKKQQHSKKRTAQADVSD 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CNLTKDPKPVSGQKCEQICNEPCEEVVLKRSSKSKRKTDKKLMKKQQHSKKRTAQADVSD 360

Qy.   361 AKLCRRKPKKVRLLSEIINANQVEDSRSDEVHRENAADPCEDDRSTIPVPMEVSMDIPVS 420
          |||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AKLCRRKPKKVRLLSEIVNANQVEDSRSDEVHRENAADPCEDDRSTIPVPMEVSMDIPVS 420

Qy    421 NHTVGEDGLKSSKNKTKRKYSDVVDDGSSLMNWLNGKKKRTGSVHHTVAHPAGNLSNKKV 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 NHTVGEDGLKSSKNKTKRKYSDVVDDGSSLMNWLNGKKKRTGSVHHTVAHPAGNLSNKKV 480

Qy    481 TPTASTQHDDENDTENGLDTNMHKTDVCQHVSEISTQRCSSKGKTAGLSKGKTHSAASTK 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TPTASTQHDDENDTENGLDTNMHKTDVCQHVSEISTQRCSSKGKTAGLSKGKTHSAASTK 540

Qy    541 YGGESTRNGQNIHVLSAEDQCQMETENSVLSHSAKVSPAEHDIQIMSDLHEQSLPKKKKK 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 YGGESTRNGQNIHVLSAEDQCQMETENSVLSHSAKVSPAEHDIQIMSDLHEQSLPKKKKK 600

Qy    601 QKLEVTREKQTMIDDIPMDIVELLAKNQHERQLMTETDCSDINRIQSKTTADDDCVIVAA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 QKLEVTREKQTMIDDIPMDIVELLAKNQHERQLMTETDCSDINRIQSKTTADDDCVIVAA 660

Qy    661 KDGSDYASSVFDTNSQQKSLASQSTQKELQGHLALTTQESPHPQNFQSTQEQQTHLRMEE 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 KDGSDYASSVFDTNSQQKSLASQSTQKELQGHLALTTQESPHPQNFQSTQEQQTHLRMEE 720

Qy    721 MVTIAASSPLFSHHDDQYIAEAPTEHWGRKDAKKLTWEQFKATTRNSPAATCGAQFRPGI 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 MVTIAASSPLFSHHDDQYIAEAPTEHWGRKDAKKLTWEQFKATTRNSPAATCGAQFRPGI 780

Qy    781 QAVDLTSTHVMGSSSNYASRQPVIAPLDRYAERAVNQVHARNFPSTIATMEASKLCDRRN 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    781 QAVDLTSTHVMGSSSNYASRQPVIAPLDRYAERAVNQVHARNFPSTIATMEASKLCDRRN 840

Qy    841 AGQVVLYPKESMPATHLLRMMDPSTLASFPNYGTSSRNQMESQLHNSQYAHNQYKGSTST 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 AGQVVLYPKESMPATHLLRMMDPSTLASFPNYGTSSRNQMESQLHNSQYAHNQYKGSTST 900

Qy    901 SYGSNLNGKIPLTFEDLSRHQLHDLHRPLRPHPRVGVLGSLLQKEIANWSENCGTQSGYK 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 SYGSNLNGKIPLTFEDLSRHQLHDLHRPLRPHPRVGVLGSLLQKEIANWSENCGTQSGYK 960

Qy    961 LGVSTGITSHQMNRKEHFEALNSGMFSAKWNALQLGSVSSSADFLSARNSIAQSWTRGKG 1020
          ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db    961 LGVSTGITSHQMNRKEHFEALNSGMFSAKWNALQLGSVSSSAYFLSARNSIAQSWTRGKG 1020

Qy   1021 KMVHPLDRFVRQDICITNKNPADFTTISNDNEYMDYR 1057
          |||||||||||||||||||||||||||||||||||||
Db   1021 KMVHPLDRFVRQDICITNKNPADFTTISNDNEYMDYR 1057


RESULT   3
Q9LYD9
ID   Q9LYD9      PRELIMINARY;      PRT;  1096 AA.
AC   Q9LYD9;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL 121.7 KDA PROTEIN (EMBRYONIC FLOWER 1).
GN   F15N18_120 OR EMF1.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Bevan M., Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
RA   Bancroft I., Mewes H.W., Rudd S., Lemcke K., Mayer K.F.X.;
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21380417; PubMed=11487698;
RA   Aubert D., Chen L., Moon Y.H., Martin D., Castle L.A., Yang C.H.,
RA   Sung Z.R.;
RT   "Emf1, a novel protein involved in the control of shoot architecture
RT   and flowering in arabidopsis.";
RL   Plant Cell 13:1865-1875(2001).
DR   EMBL; AL163815; CAB87713.1; -.
DR   EMBL; AF319968; AAK98528.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   1096 AA;  121670 MW;  190308C7A96A0F61 CRC64;


  Query Match             7.1%;  Score 393;  DB 10;  Length 1096;
  Best Local Similarity   21.0%;  Pred. No. 6.4e-16;
  Matches  250;  Conservative  189;  Mismatches  403;  Indels  348;  Gaps   58;

Qy     55 QCEHFSIRGYVALLQKKDPKFC------SLSRIFHDQKKCDEHKASSSPFSVAKFRRWDC 108
          :|:|||:||:||  :::| : |      |:| :       |:   :    || ||| | |
Db     25 KCDHFSMRGFVAETRERDLRKCWPFSEESVSLV-------DQQSYTLPTLSVPKFRWWHC 77

Qy    109 SKCLDKLKTSDNGTAPRTLPAKQNGTSDG-CSITFVRSTFVPASVGSQKVSPSTQSSQGK 167
          |:  :   :  |:    |  : | |:  :| |   ::   :    |
Db     78 MSCIKDI----DAHGPKDCGLHSNSKAIGNSSVIESKSKFNSLTIIDHEKEKKTDI---- 129

Qy    168 NADRSTLPK-SVQEGNDSKC---------NAPSGKNGAAEANTDSPMKDLQGPAQNYDVA 217
          || :   | |   || :           | |      :|  |  : |  | |   |
Db    130 -ADNAIEEKVGVNCENDDQTATTFLKKARGRPMG-----ASNVRSKSRKLVSPEQ---VG 180

Qy    218 ANVSED---NTSVDVGALPEVPQITWHIEVNGA--------DQPPSTPKLSEVVLKRNED 266
          | |::     |:|: :  |   :   :   |:        | ||   |  : :    |
Db    181 NNRSKEKLNKPSMDISSWKEKQNVDQAVTTFGSSEIAGVVEDTPPKATKNHKGIRGLMEC 240

Qy    267 ENGKTEETLVAEQCNLTKDPNPMSGKER---DQVAEQCNLTKDPKPVSGQKCEQICNEPC 323
          :|| :|   :|           ||| :|    :|      |  : |   |    :
Db    241 DNGSSESINLA----------MSGLQRRKSRKVRLLSELLGNTKTSGGSNIRK------ 283

Qy    324 EEVVLKRSSKSKRKTDKKLMKKQQH-SKKRTAQADVSD--AKLC---------------- 364
          ||  ||: |   ||  :||: :  : |:  :     |:  :| |
Db    284 EESALKKESVRGRK--RKLLPENNYVSRILSTMGATSENASKSCDSDQGNSESTDSGFDR 341

Qy    365 -----RRKPKKVRLLSEIINANQVEDSRSDEVHRENAADPCEDDRST----------IP 408
               ::: :: ::: | : :   | |:  |  :|: ||| :  |||          :|
Db    342 TPFKGKQRNRRFQVVDEFVPSLPCETSQ--EGIKEHDADPSK--RSTPAHSLFTGNDSVP 397

Qy    409 VPMEVSMDIPVSNHTVGEDGLKSSKNKTKRKYSDVVDDGSSLM----NWLNGKK--KRTG 462
          |       | :  |  |  |   | |||:    |:|:| | :    | ::| :    ||
Db    398 CP-------PGTQRT--ERKLSLPKKKTKK---PVIDNGKSTVISFSNGIDGSQVNSHTG 445

Qy    463 SVHHTVAHPAGNLSNKKVTPTASTQHDDENDTENGLDTNMHKTD--VCQHVSEISTQRCS 520
          :||:     |: |:|                 ||  |   :|    :::|::: :  :
Db    446 PSMNTVSQTRDLLNGKRV--------------GGLFDNRLASDGYFRKYLSQVNDKPIT 490

Qy    521 S---------KGKTAGLSKGKTHSAASTKYGGESTRNGQNIHVLSAEDQCQMETENSVLS 571
          |         : : |  :  :  |::|    |   | |  : ::   :       :|  :
Db    491 SLHLQDNDYVRSRDAEPNCLRDFSSSSKSSSGGWLRTG--VDIVDFRNNNHNTNRSSFSN 548

Qy    572 HSAKVSPAEHDIQIMSDL-----------------HEQSLPKKKKKQKLEVTREKQTMI 613
          :  |:  ::  :| :                     |: :   : | | |:|
Db    549 LKLRYPPSSTEVADLSRVLQKDASGADRKGKTVMVQEHHGAPRSQSHDRKETTTEEQNN- 607

Qy    614 DDIPMDIVELLAKNQHERQLM-TETDCSDINRIQSKTTADDDCVIVAAKDGSDYASSVFD 672
          |||||:||||:||||:|| |   | | |:    |       : :::   :  |   |: |
Db    608 DDIPMEIVELMAKNQYERCLPDKEEDVSNKQPSQETAHKSKNALLIDLNETYDNGISLED 667

Qy    673 TNSQQ--KSLASQSTQKELQGHLALTTQESPH----------PQN---FQSTQEQQTHLR 717
          |: :  |  :| : ::|   |  :  |:: |          |     |  ||| :
Db    668 NNTSRPPKPCSSNARREE---HFPMGRQQNSHDFFPISQPYVPSPFGIFPPTQENR---- 720

Qy    718 MEEMVTIAASSPLFSHHDDQYIAEAPTEHWGRKDAKKLTWEQFKATTRNSPAATCGA--- 774
                |||  || |: |::   ||   | ::    ::   :|       || :
Db    721 --------ASSIRFSGHNCQWLGNLPTV--GNQNPSPSSFRVLRA------CDTCQSVPN 764

Qy    775 QFRPGIQAVDLTSTHVMGSSSNYASRQPVIAPLDRYAERAVNQVHARNFPSTIATMEASK 834
          |:|         ::| :  ||       :| |  :|   ::|   : | | |::  :||
Db    765 QYR--------EASHPIWPSS-------MIPPQSQYKPVSLNINQSTN-PGTLS--QASN 806

Qy    835 LCDRRNAGQVVLYPKE--------------------------------SMPATHLLRMM 861
            :  |   |    |:                                |:|| ||| ::
Db    807 NENTWNLNFVAANGKQKCGPNPEFSFGCKHAAGVSSSSSRPIDNFSSESSIPALHLLSLL 866

Qy    862 DPSTLASFP--NYGTS--SRNQMESQLHNSQYAHNQYKGSTSTSYGSNLNGKIPLTFEDL 917
          ||   :: |   :| :  ::         : ::   |   |: ::| :    :|| |:
Db    867 DPRLRSTTPADQHGNTKFTKRHFPPANQSKEFIELQTGDSSKSAYSTK---QIP--FDLY 921

Qy    918 SRHQLHDLHRPLRP-HPRVGVLGSLLQKEIANWSENCGTQSGYKLGVSTGITSHQMNRKE 976
          |:    :  |   | · | :|   | |  : |:|| :                ||  : :
Db    922 SKRFTQEPSRKSFPITPPIGT--SSLSFQNASWSPH----------------HQEKKTK 962

Qy    977 HFEALNSGMFSAKWNALQLGSVSSSAD-----FLSARNSIAQSWTRGKGKMVHPL----- 1026
            :      |:  :|  :    :|| |      | | ||           |: ||
Db    963 RKDT-----FAPVYNTHEKPVFASSNDQAKFQLLGASNS----------MMLPLKFHMT 1006

Qy   1027 DRFVRQD---------------------ICITNKNPADFTTISNDNEYM 1054
            · |:  :|                   :|  |:||||||     | ||
Db   1007 DKEKKQKRKAESCNNNASAGPVKNSSGPIVCSVNRNPADFTIPEPGNVYM 1056


RESULT   4
Q9Y4D6
ID   Q9Y4D6      PRELIMINARY;      PRT;  1983 AA.
AC   Q9Y4D6;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
```

```
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   KIAA0624 PROTEIN (FRAGMENT).
GN   KIAA0624.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=98403880; PubMed=9734811;
RA   Ishikawa K., Nagase T., Suyama M., Miyajima N., Tanaka A., Kotani H.,
RA   Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. X.
RT   The complete sequences of 100 new cDNA clones from brain which can
RT   code for large proteins in vitro.";
RL   DNA Res. 5:169-176(1998).
DR   EMBL; AB014524; BAA31599.1; -.
FT   NON_TER       1       1
SQ   SEQUENCE   1983 AA;  221956 MW;  EE1676A3B8ED37F9 CRC64;
```

Query Match 4.0%; Score 221; DB 4; Length 1983;
Best Local Similarity 18.6%; Pred. No. 6.2e-05;
Matches 210; Conservative 171; Mismatches 423; Indels 328; Gaps 53;

```
Qy      83 HDQ---KKCDEHKASSSPFSVAKFRRWDCSKCLDKLKTSDNGTAPRTLPAKQNGTSDGCS 139
           ||:   || |    :   . | |       |::||  ::: : |           ||| |
Db     947 HDEVVDVKCHSHSPFRNERGKGKIRHH--ISCIEKLSKTESISVP---------TSDHRS 995

Qy     140 ITFVRSTFVPASVGSQKVS---------PSTQSS---QGKNADRSTLPKSVQEG------ 181
           :     : |:  : |||         |   ||    |: :    :  |:: |
Db     996 L-------IEANQSNSKVSELDTIYCTLPRKSSSFLIHGRQSGSKIMAASLRNGPPPFQI 1048

Qy     182 -------------NDSKCNAPSGKNGAAEANTDSPMKDLQGPAQNYDVAANVSEDNTSVD 228
                       |    ::|   |  ::  :||    |: |     :    |  :|||
Db    1049 KNNVEDAMGNYMLNKFSPSSPESANECSKVLSDSA---LEAPEATERMTNVKSSGSTSVR 1105

Qy     229 VGALP-------EVPQITWHIE----------VNGADQPPSTPKLSEVVLKRNEDENGKT 271
           | ||         |    |              :| |    ||:  | ::   | ::
Db    1106 KGPLPFLINRAMSCPSGEPHASTGREGRKKPLTSGMDASELTPRAWERIISPVESDSS-- 1163

Qy     272 EETLVAEQCNLTK------------------------------DPNP----MSGKERDQ 296
              |:|||                                  || |    :|||||
Db    1164 -----VRDCSLTKRQHQKENFQEYTEKEGKMAASRRSVFALSNEDPLPFCSDLSGKERGK 1218

Qy     297 VAEQCNLTKDPKPVSGQKCEQICNEPCEEVVLKRSSKSKRKTDKKLMKKQQHSKK----- 351
           :   |     ||| :     |  | |||    :  :: : |     ||:::
Db    1219 TLHKVK-TTSTFSVSGDE----DNVKCLEVVSIYYTLPRKPSKKFCNLLQQYTQNTNLLI 1273

Qy     352 RTAQADVSDAKLCRRKPKK------------VRLLSEIINANQVEDSRSDEVHRENAADP 399
           : | :          | |:                |  :|::|   : :    | :   |
Db    1274 ESPQVETETFPNALEKDKQNYSTREQSGTPSCENLKMSVNSDQTLTTENMTAFRLSNRGP 1333

Qy     400 CEDD-------RSTIPVPMEVSMDIPVSNHTVGEDGLKSSKNKTKRKYSDVVDDGSSLMN 452
                     : : :| | |    : :  : :  |  |:|| :|
Db    1334 LAPTLQEMASVEAAVSLPEEESKAREIFSDNLAKTPLGDSENKKER-------------- 1379

Qy     453 WLNGKKKRTGSVHHTVAHPAGNLSNKKVTPTASTQHDDENDTENGLDTNMHKTDVCQHVS 512
           |||  :: ::| ::     |:| :| :         :  |   : |   : :  :     :
Db    1380 ---GKKLQSETLHTSLMLQRKNVSEEK-SENCQQSINSSNSGPSSLPA-LSEVNIGNSQT 1434

Qy     513 EISTQRCSSKGKT-----AGLSKGKTHSAASTKYG--GESTRNGQNIHVLSAEDQCQMET 565
             |:  |:  |:      :|    | |:: :   |  |    | |:         ||  |
Db    1435 RRSSWECTGSGRAIPFTGSGKCPQKDHTSTAVGDGSSGSQPREGRG----DIGTNCQKMT 1490

Qy     566 ENSVLSHSAK----VSPAEHDIQIMSDLHE--------QSLPKK--KKKQKLEVTREKQT 611
           |  ||||      ::|| | :|:  :            || |::  :: |:  :|  ::
Db    1491 -NKTLSHSESQVFALTPALHKLQLGEETQSDEPNLESLQSEPRELPQRSQEANMTESRK- 1548

Qy     612 MIDDIPMDIVELLAKNQHERQLMTETDCSDINRIQSKTTADDDCVIVAAKDGSDY----- 666
                     |::: ::    :    : |:  :||  ||   :|  :: |
```

```
Db    1549 ------------AEDEMQKSAWDQPSLPEGNK--NKTNLDD---LVKGENRSSVKHRLA 1590

Qy     667 ----ASSVFDTN--SQQKSLAS---QSTQKELQGHLALTTQE------SPHPQNFQS--- 708
               ||  |      | :: :|:   ||  :    ||:| | |       | |:: :|
Db    1591 AMSKASRKFPAKDVSPRRHVATIFPQSGSRSGFDHLSLGTVECNPLFPEPTPKSAESIGE 1650

Qy     709 --TQEQQTHLRMEE--------------MVTIAASSPLFSHHDDQY--IAEAPTEHWGR 749
               | : |::  |                |:   |: :   | :::  ::|:|::|
Db    1651 SRLSENRKHVKKSENLLPITVLPNREPSTHVSNQKSNSISQRHQNEFKNVSESPSKHENS 1710

Qy     750 KDAKKLTWEQFKATTRNSPAATCGAQFRPGIQAVDLTSTHVMGSSSNYASRQPVIAPLDR 809
           ||          |   :    ||        :  ||      | |     |   ||:
Db    1711 KD----------VTAAQNLVRESGAP-----SPITFTSLREAEFSDNQRRLSPPF-PLEP 1754

Qy     810 YAERAVNQVHARNFPSTIATMEASKLCDRRNAGQVVLYPKESMPATHLLRMMDPSTLASF 869
            |::           | :::  || |  :|:|  :     |  |  || |     :| |
Db    1755 -AQK----------SRVSSPLASFLQQQRSASSL-----EWEPEPHLYR---SKSLKSI 1794

Qy     870 PNYGTSSRNQMESQLHNSQYAHNQYKGSTSTSYGSNLNGKIPLTFEDLSRHQLHDLHRPL 929
              :|   |       |  :      :  |||     |   :: |  |   :      |
Db    1795 NVHGDLLRKS-----HPPKVRERHFSESTSI---DNALSRLTLGNEFSVNNGYSRRFRSF 1846

Qy     930 RPHPRVGVLGSLLQKEIANWSENCGTQSGYKLGVSTGITSHQMNRKEHFEALNSGMFSAK 989
              |     |:       :|:   ||::| : :|          :  :: |:|  :
Db    1847 SELPSCD--GN------ESWAYRSGTKTGPRSAISI------------YRPIDYGIFGKE 1886

Qy     990 WNALQLGSVSSSADFLSARNSIAQSWTRGKGKMVHPLDRFVRQDICITNKNP 1041
             | :|  |   |:       |: :  |        |:: |:    :||
Db    1887 QQLAFLENVKRS---LTQGRLWKPSFLKNPG--------FLKDDL----RNP 1923
```

```
RESULT   5
Q9LH98
ID   Q9LH98      PRELIMINARY;      PRT;  2081 AA.
AC   Q9LH98;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   GENOMIC DNA, CHROMOSOME 3, BAC CLONE: T19N8.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COLUMBIA;
RA   Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL   Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COLUMBIA;
RX   MEDLINE=20363099; PubMed=10907853;
RA   Nakamura Y.;
RT   "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT   Sequence features of the regions of 4,251,695 bp covered by ninety P1,
RT   TAC and BAC clones.";
RL   DNA Res. 7:217-221(2000).
DR   EMBL; AP002057; BAB03174.1; -.
SQ   SEQUENCE   2081 AA;  232851 MW;  D3603E1F85EFFF29 CRC64;
```

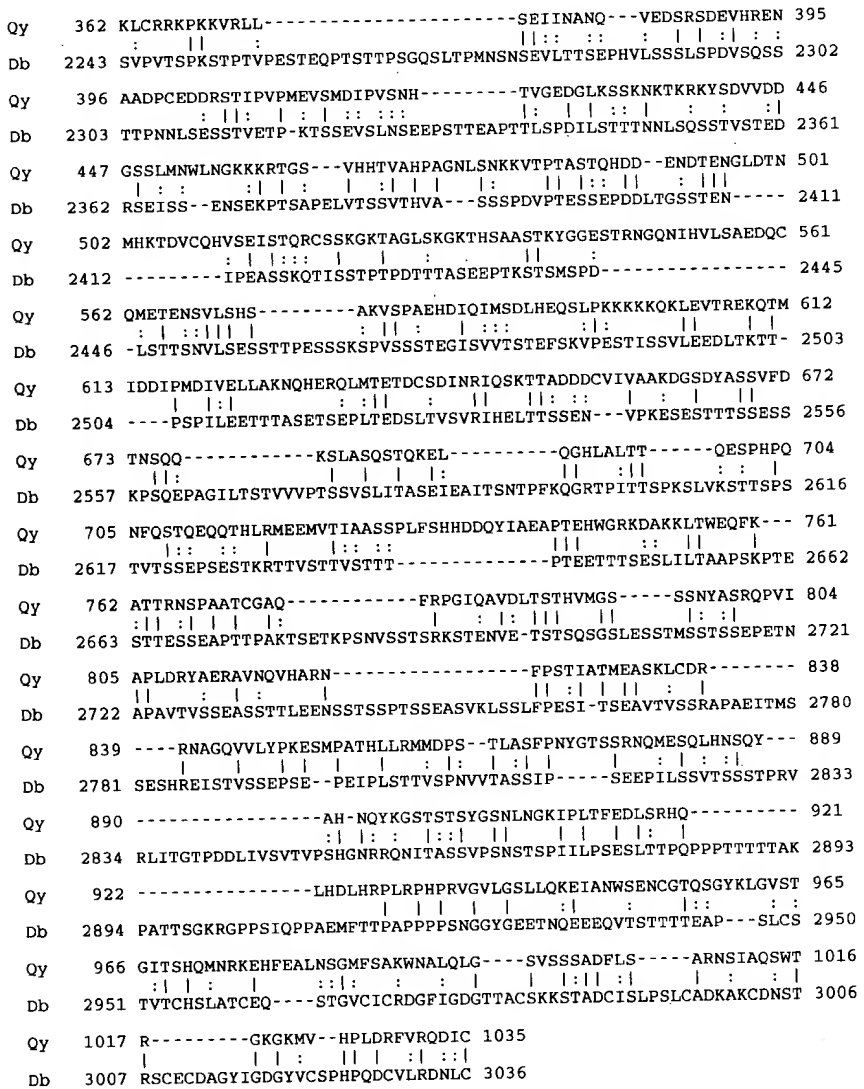
Query Match 3.9%; Score 216; DB 10; Length 2081;
Best Local Similarity 18.9%; Pred. No. 0.00013;
Matches 155; Conservative 139; Mismatches 382; Indels 144; Gaps 26;

```
Qy     153 GSQKVSPSTQSSQGKNADRSTLPKSVQEGNDSKCNAPSGKNGAAEANTD------SPMKD 206
           |::| |   :  : | :  :   |||:  :: | ::   :: | | : :          ||
Db     777 GNKKESEKVEKGEKKESKDA---KSVETKDNKKLSSTENRDEAKERSGEDNKEDKEESKD 833

Qy     207 LQG-----PAQNYDVAANVS--------EDNTSVDVGALPEVPQITWHIEVNGADQPPST 253
           |        :|  |  ||         :|: ||:| |  |        ||   |: ||
Db     834 YQSVEAKEKNENGGVDTNVGNKEDSKDLKDDRSVEVKANKEESMKKKREEVQRNDK-SST 892
```

```
Qy    254 PKLSE----------------VVLKRNEDENGKTEETLVAEQCNLTKDPNPMSGKERDQV 297
          :: :                  |  |::| : |  ||    :  | :         |:: :
Db    893 KEVRDFANNMDIDVQKGSGESVKYKKDEKKEGNKEEN--KDTINTSSKQKGKDKKKKKKE 950

Qy    298 AEQCNLTKDPKPVSGQKCEQICNEPCEEVVLKRSSKSKRKTDK----KLMKKQQHSKKRT 353
          ::  |: |  :      | | : ||       ||:   :|::| |    || :: : :|::
Db    951 SKNSNMKKKEE----DKKEYVNNE------LKKQEDNKKETTKSENSKLKEENKDNKEKK 1000

Qy    354 AQADVSDAKLCRRKPKKVRLLSEIINANQVEDSRSDEVHRENAADPCEDDRSTIPVPMEV 413
             |   |   | | :     |:     : |  :| :  ||      : :         |
Db   1001 ESED--SASKNREKKEYEEKKSKTKEEAKKEKKKSQDKKREEK----DSEERKSKKEKEE 1054

Qy    414 SMDIPVSNHTVGEDGLKSSKNKTKRKYSDVV--DDGSSLMNWLNGKKKRTGSVHHTVAHP 471
          | |:             | |:|   :|  |    :|  |:    : |:|:         |
Db   1055 SRDLKAKKKEEETKEKKESENHKSKKKEDKKEHEDNKSMKKEEDKKEKK--------KHE 1106

Qy    472 AGNLSNKKVTPTASTQHDDENDTENGLDTNMHKTDVCQHVSEISTQRCSSKGKTAGLSKG 531
               |:       : :|:|  :   | |  |    ||| ::  :    | |       |
Db   1107 ESKSRKKEEDKKDMEKLEDQNSNKKKEDKNEKKKS--QHV-KLVKKESDKKEKKENEEKS 1163

Qy    532 KTHSAASTKYGGESTRNG-QNIHVLSAEDQCQMETENSVLSHSAKVSPAEHDIQIMSDLH 590
          :|    |:|    | :|        |::|| : : :    |   |:   | | :  : :
Db   1164 ETKEIESSK----SQKNEVDKKEKKSSKDQQKKKEKEMKESEEKKLKKNEEDRKKQTSVE 1219

Qy    591 EQSLPKKKKKQKLEVTREKQTMIDDI-----PMDIVELLAKNQHERQLMTETDCSDINR- 644
          |    |: ||:| :   :|:             |:     |:|| : |  |: |  :
Db   1220 ENKKQKETKKEKNKPKDDKKNTTKQSGGKKESMESESKEAENQQKSQATTQADSDESKNE 1279

Qy    645 --IQSKTTADDDCVIVAAKDGS--------DYASSVFDTNSQ----QKSLASQSTQKELQ 690
            :|: : ||      |  | |         |  ::    | :    | |:|    |||
Db   1280 ILMQADSQADSHSDSQADSDESKNEILMQADSQATTQRNNEEDRKKQTSVAENKKQKE-- 1337

Qy    691 GHLALTTQESPHPQNFQSTQEQQTHLRMEEMVTIAASSPLFSHHDDQYIAEAPTEHWGRK 750
              | :|   |:: :    :|:  : | | : :  :      ::|  ::| |:    :
Db   1338 -----TKEEKNKPKDDKKNTTKQSGGKKESMESESKEA------ENQQKSQATTQADSDE 1386

Qy    751 DAKKLTWEQFKATTRNSPAATCGAQFRPGIQAVDLTSTHVMGSSSNYASRQPVIAPLDRY 810
             ::                        :||     :|    : :  |:  ::   |
Db   1387 SKNEIL----------------------MQADSQADSHSDSQADSDESKNEILMQADSQ 1423

Qy    811 AERAVNQVHARNFPSTIATMEASKLC---------DRRNAGQVVLYPKESMPATHLLRMM 861
          |   |   |  :::|  :  |             |::|  :     |||| :
Db   1424 ATTQRNNEEDRKKQTSVAENKKQKETKEEKNKPKDDKKNTTEQSGGKKESMESESKEAEN 1483

Qy    862 DPSTLASFPNYGTSSRNQMESQLHNSQYAHNQYKGSTSTS 901
             : |:       |:|::  |  :    |  :| :  |
Db   1484 QQKSQATTQGESDESKNEILMQADSQADTHANSQGDSDES 1523
```

```
RESULT   6
Q23587
ID   Q23587      PRELIMINARY;      PRT;  3507 AA.
AC   Q23587;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   ZK783.1 PROTEIN.
GN   ZK783.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Favello A., Vaudin M.;
RT   "The sequence of C. elegans cosmid ZK783.";
RL   Submitted (NOV-1994) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U13646; AAC24418.1; -.
DR   HSSP; P00736; 1APQ.
DR   InterPro; IPR000152; Asx_hydroxyl.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR001254; Trypsin.
DR   InterPro; IPR001507; zona_pellucida.
DR   Pfam; PF00008; EGF; 15.
DR   SMART; SM00179; EGF_CA; 12.
DR   SMART; SM00001; EGF_like; 16.
DR   SMART; SM00241; ZP; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 9.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR   PROSITE; PS01186; EGF_2; 22.
DR   PROSITE; PS01187; EGF_CA; 13.
DR   PROSITE; PS00135; TRYPSIN_SER; UNKNOWN_1.
KW   Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
SQ   SEQUENCE   3507 AA;  365315 MW;  154F0B687874D9DF CRC64;
```

```
  Query Match             3.9%;  Score 215.5;  DB 5;  Length 3507;
  Best Local Similarity   17.9%;  Pred. No. 0.00029;
  Matches  242;  Conservative  179;  Mismatches  510;  Indels  419;  Gaps   52;

Qy     12 RVVGTNCMLARGGTGAVAPVLELTATPRQDAAA-------------EAGVDEPAQHQC- 56
          |:    |    | ||  |  :::   |: |  |                :  |||    ||
Db   1780 RMGSCGCKCMAGYTGDGATCIKIEEEPKSDKTACTDEWSRLCELEKKQCTVDEEEVPQCG 1839

Qy     57 ----EHFSIRGYVALLQ-------------------KKDPKFCSLSRIF---------- 82
               |  | |    ||                        |  |||    |
Db   1840 ACLPGHHPINGTCQSLQISGLCAQKNDCNKHAECIDIHPDSHFCSCPDGFIGDGMICDDV 1899

Qy     83 ----------HDQKKCDEHKASSSPFSVAKFRRWDCSKCL---------DKLK------- 116
                    :  ||:    | :   :  |:: |  ||:              :|::
Db   1900 DECNNAGMCDDENTKCENTIGSFNCVCLEGFKKVD-EKCVVDEKKQPNREKIEIDEENSS 1958

Qy    117 --------------TSDNGTAPRTLPAKQNGTSDGCSITFVR---STFVPASVGSQKVS 158
                         :| : |:  :  |: : |:   | |  :   |:  | :|     |
Db   1959 SSNSGQEKPTTKGIVSSTSATSSESTTAEPHVTTSISSTTSTKDMTSSKSPENVTMSSES 2018

Qy    159 P--STQSSQGKNADRSTLPKSVQEGNDSKCNAPSGKNGAAEANTDSPMKDLQGPAQNYDV 216
          |  || ||:   |  :|:  :  | : |:    |       |  |:| :|
Db   2019 PEVSTSSSKSTTASETTVSSTPSESSSSEAPLTSSPATTTEVITESSVKS---------- 2068

Qy    217 AANVSEDNTSVDVGALPEVPQITWHIEVNGADQPPSTPKLSEVVLKRNEDENGKTEETLV 276
             | :: : |    : |::|     :      ||||  :   :    |  |:  |::
Db   2069 TTPKEESSSEITVKLSSKSPEVT----ESSVKSSPSTPSTTSQSVTSTVPETSKS--TVL 2122

Qy    277 AEQCNLTK---------------DPNPMSGKERDQVAEQCNL-----TKDPK------- 308
          : :  :|                 : :|          :|     |  |:
Db   2123 SSEAPVTSTSPTEVHTSSETKPSLSASSTTGDTNSTTPSTSSLASVKSTSAPEGTSASVA 2182

Qy    309 PVSGQKCEQICNEP-------CEEVVLKRSSKSKRKTDKKLMKKQQHSKKRTAQADVSDA 361
          ||          ::|        |   :: | |   : |   : : |  | :  :    :
Db   2183 PVKLSSLSPDVSQPSTKTFDATESSTVQASETSSGTSVKSTSEPESHVTKLSITSSNPSS 2242
```



RESULT 7
O44929
ID O44929 PRELIMINARY; PRT; 1690 AA.
AC O44929;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE MICROTUBULE BINDING PROTEIN D-CLIP-190.
GN CLIP-190 OR CG5020.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OREGON R;
RA Lantz V.A., Miller K.G.;
RL J. Cell Biol. 0:0-0(1998).
DR EMBL; AF041382; AAB96783.1; -.
DR FlyBase; FBgn0020503; CLIP-190.
DR InterPro; IPR000938; CAP-Gly.
DR Pfam; PF01302; CAP_GLY; 2.
SQ SEQUENCE 1690 AA; 189103 MW; BE4F48FD15F17A7C CRC64;

Query Match 3.9%; Score 215; DB 5; Length 1690;
Best Local Similarity 19.1%; Pred. No. 0.00012;
Matches 171; Conservative 139; Mismatches 352; Indels 234; Gaps 33;

```
Db   1044 AEGERKEKS---------FEESIKN--------------------LQEEVTKAKTENLE 1073

Qy    951 ENCGTQSGYK-LGVSTGITSHQMNRKEHF---EALNSGMFSAKWNALQLGSVSSSA 1002
          : |||:  | |     ||: ::  ||     :|          |:|: : : ||
Db   1074 LSTGTQTTIKDLQERLEITNAELQHKEKMASEDAQKIADLKTLVEAIQVANANISA 1129


RESULT   8
Q9SZ55
ID   Q9SZ55      PRELIMINARY;      PRT;   852 AA.
AC   Q9SZ55;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   HYPOTHETICAL 92.0 KDA PROTEIN.
GN   F11C18.80 OR AT4G31880.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Bevan M., Terryn N., Ardiles W., Buysshaert C., Dasseville R.,
RA   De Clerck R., De Keyser A., Neyt P., Rouze P., Van Den Daele H.,
RA   Villaroel R., Gielen J., Van Montagu M., Hoheisel J., Mewes H.W.,
RA   Mayer K.F.X., Schueller C.;
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Terryn N., Ardiles W., Buysshaert C., Dasseville R., De Clerck R.,
RA   De Keyser A., Neyt P., Rouze P., Van Den Daele H., Villaroel R.,
RA   Gielen J., Van Montagu M., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL049607; CAB40758.1; -.
DR   EMBL; AL161579; CAB79906.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   852 AA;  91950 MW;  AEA128F4D5A0F59D CRC64;


  Query Match             3.8%;  Score 209.5;  DB 10;  Length 852;
  Best Local Similarity   21.2%;  Pred. No. 0.0001;
  Matches  174;  Conservative   97;  Mismatches  282;  Indels  269;  Gaps   39;

Qy     31 VLELTATPRQ-----DAAAEAGVDEPAQHQCEHFSIRGY-------------VALLQKK 71
          :||  |  |       |   :| : |  ||  :  :|| :             : | : :
Db    129 ILETVAKVRSCVVMLDLECDALLIEMFQHFLK--AIRDHHSGNVFSSMENIMTLVLEESE 186

Qy     72 DPKFCSLSRIFHDQKKCDEHKASSSPFSVAKFRRWDCSKCLDKLKT------SDNGTAPR 125
          |     || | |  || ||    |      :       | |  |'||        :|
Db    187 DIPSEMLSPILHSVKKDDEISQVS-----RRLAEQVLSNCASKLKTYLTEAVKSSGV--- 238

Qy    126 TLPAKQNGTSDGCSITFVRSTFVPASVGSQKVSPSTQSSQGKNADRSTLPKSVQEGNDSK 185
          |    |  :  |  ||       |    | |:   : |||     : : |: :     :
Db    239 PLDKYSNIVASICEGTF------SALQQDQVVANEKEDSQGHIKRETEVEKAAEISTPER 292

Qy    186 CNAP---SGK----NGAAEAN-----TDSPMKD----LQGPAQNYDVAANVSEDNTSVDV 229
           :||   |||    || |: |     |||  |      :   |  |   |   :||:
Db    293 TDAPKDESGKSGVSNGVAQQNDSSVDTDSMKKQDDTGAKDEPQQLDNPRNTDLNNTT--- 349

Qy    230 GALPEVPQITWHIE--------VNGAD-----------------------QPPSTPKLS 257
            | | :   ||         |  ||                        ||    ::
Db    350 ---EEKPDVEHQIEEKENESSSVKQADLSKDSDIKEETEPAELLDSKDVLTSPPVDSSVT 406

Qy    258 EVVLKRNE--------------DENGKTEETLVAEQCNLTKDPNPMSGKERDQVAEQCNL 303
                ||              ||        :||:      |   : ::: :     :
Db    407 AATSSENEKNKSVQILPSKTSGDETANVSSPSMAEELPEQSVPKKTANQKKKE-----SS 461

Qy    304 TKDPKPVSGQKCEQICNEP--CEEVVLKRSSKSKRKTDKKLMKKQQHSKKRTAQADVSDA 361
          |:: || :    |::  ||   |  | |:| | |  :  |       ||| |     |:
Db    462 TEEVKPSASIATEEVSEEPNTSEPQVTKKSGK-KVASSSKTKPTVPPSKKST-----SET 515

Qy    362 KLCRRKPKKVRLLSEIINANQVEDSRSDEVHRENAADPCEDDRSTIPVPMEVSMDIPVSN 421
          |: ::  |||     : : |  | ::  |  ::       |: |
Db    516 KVAKQSEKKV-----VGSDNAQESTKPKEEKKKPGRGKAIDEESL-------------- 555

Qy    422 HTVGED--------GLKSSKNK--------------TKRKYS----------DVVDDGSSL 450
          ||   |        |  :||:|              |||| |          :|  || :
Db    556 HTSSGDNEKPAVSSGKLASKSKKEAKQTVEESPNSNTKRKRSLGQGKASGESLV--GSRI 613

Qy    451 MNW-------------LNGKKKRTGSVHHTVAHPAGN-----LSNKKVTPTASTQ--HD 489
           |               :  ||:     | | :  |:     | |:| :|  .::   |
Db    614 KVWWPMDQAYYKGVVESYDAAKKK-----HLVIYDDGDQEILYLKNQKWSPLDESELSQD 668

Qy    490 DENDTENGLDTNMHKTDVCQHVSEISTQRCSSKGKTAGLSKGKTHS-----AASTKYGGE 544
          :|   : | :          :  | : :   ||| |    ||    |     |: :|   |
Db    669 EEAADQTGQE---------EDASTVGSGAGSSKAKATPASKSSKTSQDDKTASKSKDSKE 719

Qy    545 STRNGQNIHVLSAEDQCQMETENSVLSHSAKVSPAEHDIQIMSDLHEQSLPKKKKKQKLE 604
          ::|  :     |:|:: : |      :  |   | :: ||  :|   :     |||::  :
Db    720 ASREEE----ASSEEESEEEEPPKTVGKSGS-SRSKKDISSVSKSGKSKASSKKKEEPSK 774

Qy    605 VT---REKQTMIDDIPMDIVELLAKNQHERQLMTETDCSDINRIQSKTTADDDCVIVAAK 661
           |   : |   :  :|                                :|||          ||
Db    775 ATTSSKSKSGPVKSVP--------------------------AKSKTGKG------KAK 801

Qy    662 DGSDYASSVFDTNSQQKSLASQSTQKELQGHLALTTQESPHP 703
          ||  ||:     |: |  ||:|  :|          | : | |
Db    802 SGS--AST---PASKAKESASESESEE--------TPKEPEP 830


RESULT   9
Q22257
ID   Q22257      PRELIMINARY;      PRT;  1295 AA.
AC   Q22257;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   T06E4.1 PROTEIN.
GN   T06E4.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Lloyd C.R.;
RL   Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
DR   EMBL; Z70756; CAA94789.1; -.
SQ   SEQUENCE   1295 AA;  147395 MW;  57C70DBCCD2172F8 CRC64;


  Query Match             3.8%;  Score 208.5;  DB 5;  Length 1295;
  Best Local Similarity   18.4%;  Pred. No. 0.0002;
  Matches  208;  Conservative  179;  Mismatches  402;  Indels  339;  Gaps   52;

Qy     84 DQKKCDEHKASSSPFSVAKFRRWDCSKCLDKLKT-SDNGTAPRTLPAKQNGTSDGCSITF 142
          |   |:  :|||:          |    |:: |  |:: :     | ::  :||  :|
Db     96 DNSVCEPPRASSA---------HDVDPNLEQTKVYSEDSISKDCSIADESFATDG-DVTL 145

Qy    143 VRSTFVPASVGSQKVSPSTQSSQGKNADRSTLPKSVQEGNDSKCNAPSGKNG----AAEA 198
```

```
Db    146 ANDFFQQATIQDLENTPST----------PLRSQTQVGDESTPTAPSPLRSKIPIESEV 194

Qy    199 NTDSPMKDLQGPAQN---------YD---VAANVSEDNTSVD----VGALPEVPQITWHI 242

Db    195 VQSTSVESVQSTSQHEKSRIEYSLYEFEMLEKNIEHDEVVKDLQLKIELLGEKHRQTL-V 253

Qy    243 EVNGA---DQPPSTPKLSEVVLK-RNEDENGKTEETLVAE-----QCNLTKDPNPMSGKE 293

Db    254 EIKEAREVEEKMLMQQVEEAVKKTKSEREAAKTLEQLLEKRIEELEMKLT-EPN----GE 308

Qy    294 RDQVAEQCNLTKDPKPVSGQKCEQICNEPCEEVVLKRSSK--SKRKTDKKLMKKQQHSKK 351

Db    309 KLQFEHQLEELK-------SRCEELTDKALKVDVMQHSIEDYEKKFVELQEMKEEADEQL 361

Qy    352 RTAQADVSDAKLCRRKPKKVRLLS----EIINANQVEDSRS----------DEVHR-ENA 396

Db    362 QKAKEDIETLQM-----KYVELETTINKEVFSNSEIETLKSEHEIVRKLMLDEIHRLENE 416

Qy    397 ADPCEDDRSTIPV-PMEVSMD---IPVSNHTVGEDGLKSS----KNKTKRKYSDVVDDGS 448

Db    417 MSALQPKNDTTELEELQKTLDDLKIDCCNLTSKMLELQSELVEVKEKATSEIGEAVQKNG 476

Qy    449 SLMNWLNGKKKRTGSV---------HHTVAHPAGNLSNKKVTPTASTQHD---------D 490

Db    477 ELLEQINSLRVENAKLVDMEGQLNDAHRKAEDKDVRISELLTTIESLRQDSEASDKLLMD 536

Qy    491 ENDTENGLDTNMHKTDVCQHVSEISTQRCSSKGKTAGLSKGKTHSAASTKYGGESTRNGQ 550

Db    537 SESTQNEYSLALENT-----VSELETMRREYKA-------------------------- 564

Qy    551 NIHVLSAEDQCQMETENSVLSHSAKVSPAEHDIQIMS-DLHEQSLPKKKKKQ--KLEVTR 607

Db    565 -----SVDKVCSLQLELEEIQHETSVELEEAEIRIKELELAQEEAVKTGSSQLKKLEIVQ 619

Qy    608 EK-QTMIDDIPMDIVELL--------------AKNQHERQ------LMTETDCSDIN--- 643

Db    620 EDCQKLRDQLKEEQIQQLVSLRETSEVMHQESARHQEEKYQIQSKLMSTEAEVIELRSSI 679

Qy    644 -------RIQSKTTADDDCVI------------VAAKDGSDYASS---VFDTNSQQKSL- 680

Db    680 DSLQAEVRVQSDSAADQKHILEDYLRKIRQAEETNEKLRSDLASSEEQILDLKNQQESLI 739

Qy    681 --------ASQSTQKELQGHLALTTQESPHPQNFQSTQEQQTHLRMEEMVTIAASSPLFS 732

Db    740 DDLKEKLHSAESTNQELQVSLEMLKIEVSNAR--QKVMESEVLKESFEALQLELSA---S 794

Qy    733 HHDDQYIAEAPTEHWGRKDAKKLTWEQFKATTRNSPAATCGAQFRPGIQAVDLTSTHVMG 792

Db    795 QEVSRSVVDAAVQ---EKDGLLRLVDTLKLKIEDTEKS----------AQDLQQSSV-- 838

Qy    793 SSSNYASRQPVIAPLDRYAERAVNQVHARNFPSTIATMEASKLCDRRNAGQVVLYPKESM 852

Db    839 -----------------EEIKQLQLDLQNFKQNAEVLESL--------------NEKL 865

Qy    853 PATHLLRMMDPSTLASFPNYGTSSRNQMESQLHNSQYAHNQYKGSTSTSYGSNLNGKIPL 912

Db    866 NSSH---KRDMVALAS----------QLEELQHKLVVGESQVENVKEELIGAKI------ 906

Qy    913 TFEDLSRHQLHDLHRPLRPHPRVGVLGSLLQKEIANWSENCGTQSGYKLGVSTGITSHQM 972

Db    907 ----MNKEMVDELN---------AKLGDALE----------GME---ELKKSLEVSEAKV 940

Qy    973 NRKEH------------------FEALNSGMFSAKWNALQLGSVSSSADFLSARNSIAQ 1013

Db    941 QRREEELIAQVSKHRDQQEQLQLTLDELKSAQHSTETSRSQSNELAARIEELEASISFAQ 1000

Qy   1014 SWTRGKGKMVHPLDRFVRQDICITNKNPA------DFTT--ISNDNEY 1053

Db   1001 KALQDVEDVKH------QQDIQISEANEAMVKLKQDFETERTSLQNEF 1042

RESULT 10
```

```
Q9QJ16
ID   Q9QJ16      PRELIMINARY;      PRT;  1520 AA.
AC   Q9QJ16;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   U86.
GN   U86.
OS   Human herpesvirus 6B.
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Betaherpesvirinae; Roseolovirus.
OX   NCBI_TaxID=32604;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Z29;
RX   MEDLINE=99412318; PubMed=10482553;
RA   Dominguez G., Dambaugh T.R., Stamey F.R., Dewhurst S., Inoue N.,
RA   Pellett P.E.;
RT   "Human herpesvirus 6B genome sequence: coding content and comparison
RT   with human herpesvirus 6A.";
RL   J. Virol. 73:8040-8052(1999).
DR   EMBL; AF157706; AAD49674.1; -.
SQ   SEQUENCE   1520 AA;  171362 MW;  0F498EEDCDF70F3D CRC64;
```

Query Match 3.8%; Score 207.5; DB 12; Length 1520;
Best Local Similarity 18.6%; Pred. No. 0.00029;
Matches 200; Conservative 170; Mismatches 480; Indels 227; Gaps 41;

```
Qy     84 DQKKCDEHKASSSPFSVAKFRRWDCSKCLDKL-KTSDNGTAPRTLPAK---QNGTSDGCS 139

Db    270 DNKEHKKQQVSTSPDAISLSYR-PSSQKMDLIKKIYDTEVIP--LPKEALIDNGSYCGVD 326

Qy    140 ITFVRSTFVPASVGSQKVSPSTQSSQGKNADRSTLPK-SVQEGNDSKCNAPSGKNGAAEA 198

Db    327 TQKYKKTHIRCR--------SIQKTKGHSS--QTINKHKVQKHNEN--HVPS-------- 366

Qy    199 NTDSPMKDLQGPAQNYDVAANVSEDNTSVD--VGALPEVPQITWHIEVNGADQPPSTPKL 256

Db    367 --RSDLKQRKS-NQHEDEAVTEARDFSKLDPLLSPLPMTPE---------------PTL 407

Qy    257 SEVVLK---RNEDENGKTEETLVAEQCNLTKDPNPMSGKERDQVAEQCNLTK---DPKPV 310

Db    408 NFAVHKTKIHSDSELHHTKKNIHRSKTSL--QDRVLISKHAPRAPTKDNSYKKHHDPKDT 465

Qy    311 SGQKCEQICNEPCEEVVLKRSSKSKRKTDKKLMKKQQHSKKRTAQADVSDAKLCRRKPKK 370

Db    466 NDPK-------------MKHSRGRTTSKKNTANSNGHQDVREVAVKNVSGKAAKNTSKK 511

Qy    371 VRLLSEIINANQVEDSRSDEVHRENAADPCEDDRSTIPVPMEVSMDIPVSNHTVGEDGLK 430

Db    512 ----SD--PSHNLHGKTSDEQYKTS------PDNEKISTPPKSKTHHCIHDSSSSEEGQY 559

Qy    431 SSKNKTKRKYSDVVDDGSSLMNWLNGKKKRTGSVHHTVAHPAGN-LSNK------KVTPT 483

Db    560 KSPNNSESNY-------------------------------GNCLSDKFYDNFTKTTPN 587

Qy    484 ASTQHDDENDTENG-----LDTNMHKTDV-------------------------CQHVS 512

Db    588 SKTNHKTEESTENTDLNSFSNENTNKTEIEDSNIIQPFSQLFCNETIIPSTSACPTQETP 647

Qy    513 EISTQRCSSKGKTAGLSKG-------KTHS------AASTKYGGESTRNGQNIHVLSAED 559

Db    648 STMNRNCASKKHSAGANKNLTDNSPIRSHSNPSSFTAFNKSNSGNSTMNS-----TSNGD 702

Qy    560 QCQMETENSVLSHSAKVSPAEHDIQIMSDLHEQSLPKKKKKQKLEVTREKQTMIDDIPMD 619

Db    703 ECTDKKPN--------CSSTENKSETSNQTNGENSDKPLSKTFTEVSDRASSRASSRASS 754

Qy    620 IVELLAKNQHERQLMTETDCSDINRIQSK--TTADDDCVIVAAKDGSDYASSVFDTNSQQ 677

Db    755 RASSRASSRASSRASSRASSRTSSRASSRAFSRASSRVSSRASSRASSRASSRASSRASS 814
```

```
Qy     678 KSLASQSTQKELQGHLALTTQESPHPQNFQSTQ-EQQTHLRMEEMVTIAASSPLFSHHDD 736
Db     815 RASGRASGRASSRASSRVSSRASSRASSRASSRASSRASSRASSRVSSKASSRASSRASS 874

Qy     737 QYIAEAPTEHWGRKDAKKLTWEQFKATTRNSPAATCGAQFRPGIQAVDLTSTHVMGSSSN 796
Db     875 RASSRASSRASSRASSRASSRASSRASSRASSRASSRASSRASSRASSRASSRASSRASS 934

Qy     797 YASRQPVIAPLDRYAERAVNQVHARNFPSTIATMEA-SKLCDRRNAGQVV-------LYP 848
Db     935 RASGRASSKASSRASSRASGKTPNNKLMSNIPSPQTYEKTSNKRKPRQIYCDSNKRQIYP 994

Qy     849 KESMPATHLLRMMDPSTLASFPNYGTSSRNQMESQLHNSQYAHNQY--------KGSTST 900
Db     995 HDTSISTEAEVSEIKFRCGPELNFYKNAAARLQSFNHNDQFYNPRFRPHIRTNRKKSEST 1054

Qy     901 SYGSNLNGKIPLTFEDLSRHQLHDLHRPLRPHPRVGVLGSLLQKEIANWSENCGTQSGYK 960
Db    1055 NYTDSESS--TSRSRSHSRYSPDSLNTPKRKKHKSG--SSSISSSI---EENSRSNSRTE 1107

Qy     961 LGVSTGIT---SHQMNRKEHFEALNSGMFSAKWNALQLGSVSSSAD--FLSARNSIAQ-- 1013
Db    1108 TGTDKLITFNYQHSRTRSSSSSSVSSSCSSSRSKSWRKSRYKNSEDPATYSPRSRLQQLE 1167

Qy    1014 ------SWTRGKGKMVHP--------LDRFVRQDICITNKNPADFTTISNDNEYMDY 1056
Db    1168 KQTRPKSASRDKTKIKSPNHESKHRHADMF-RNSQKTGEKFPLDNSSPSNTHEQSNH 1223
```

```
RESULT  11
P91257
ID   P91257      PRELIMINARY;      PRT;  3484 AA.
AC   P91257;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL 385.7 KDA PROTEIN.
GN   F12F3.3.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   None;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Fulton B., Wohldmann P.;
RT   "The sequence of C. elegans cosmid F12F3.";
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RT   "Direct Submission.";
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U80022; AAC25885.2; -.
DR   InterPro; IPR001064; Crystallin.
DR   InterPro; IPR003962; FnIII_repeat.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR000130; Zn_MTpeptdse.
DR   Pfam; PF00041; fn3; 7.
DR   Pfam; PF00047; ig; 10.
DR   PRINTS; PR00014; FNTYPEIII.
DR   SMART; SM00060; FN3; 8.
DR   SMART; SM00409; IG; 12.
DR   SMART; SM00408; IGc2; 8.
DR   SMART; SM00410; IG_like; 1.
DR   PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
DR   PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
KW   Hypothetical protein; Immunoglobulin domain; Repeat.
SQ   SEQUENCE   3484 AA;  385746 MW;  21DD6F6893020F4F CRC64;
```

```
  Query Match             3.7%;  Score 206;  DB 5;  Length 3484;
  Best Local Similarity   18.5%;  Pred. No. 0.0011;
  Matches  214;  Conservative  161;  Mismatches  419;  Indels  360;  Gaps   46;

Qy      31 VLELTATPRQDAAAEAGVDEPAQHQCEHFSIRGYVALLQ---KKDP---KFCSLSRIFHD 84
Db     368 IIKKKETPESEQVTAA---EPEQQKISEVDVQS-VAETEVGAKKKPDAEKPTDLSKAKKD 423

Qy      85 --QKKCDEHKASSSPFSVAKFRRWDCSKCLDKLKTSDNGTAPRTLPAKQNGT-------- 134
Db     424 SKSKKSDEPEASTEEKSTTEKPTND--------KTSKKSAEKKTVKPKKEVTGKPLEAKK 475

Qy     135 --------------------SDGCSITFV-RSTFVPASVGSQKVSPSTQSSQ-------- 165
Db     476 PVEDKKDASQPSSSKESSPPTDGKKKKQIPKALFIPDEISSRFGDPSTMHSETNITTTIR 535

Qy     166 GK--NADRST-----LPKSV-----------------QEGNDSKCNAPSG--------- 191
Db     536 GREGSADAKTPLVEPLSASVSMKVESAKEKAEFSFKRRSETPDDKSRKKEGLPPAKKSEK 595

Qy     192 ------KNGAAEANTDSPMKDL-------QGPA-QNYDVAANVSE----DNTSVDVGALP 233
Db     596 KDEVTAEKQSTEALIESKKKEVDESKISEQQPSDKNKSEVVGVPEKAAGPETKKDVSEIE 655

Qy     234 EVP-QITWHIEVNGADQPPSTPKLSEVVLKRNEDENGKT-----------EETLVAEQC 280
Db     656 EVPKKKTIKKKTEKSD---SSISQKSNVLKPADDDKSKSDDVTDKSKKTTEDQTKVATDS 712

Qy     281 NLTKDPNPMSGKERDQVAEQCNLTKDPKPVSGQKCEQICN--------------EPCEE 325
Db     713 KLEKAADTTKQIETETVVDD----KSKKKVLKKKTEKSDSFISQKSETPPVVEPTKPAES 768

Qy     326 VVLK----RSSKSKRKTDKKLMKKQQHSKKRTAQ-------------ADVSDAKLCRRK 367
Db     769 EAQKIAEVNKAKKQKEVDDNLKREAEVAAKKIADEKLKIEAEANIKKTAEVEAAKKQKEK 828

Qy     368 PKKVRLLSEIINANQVED------------SRSDEVHRE---NAADPCEDDRSTIPVPM 411
Db     829 DEQLKLETEVVSKKSAAEKLELEKQAQIKKAAEADAVKKQKELNEKNKLEAAKKSAADKL 888

Qy     412 EVSMDIPVSNHTVGEDGLKSSKNK--------------TKRKYSDVVDDGSSLMNWLNG 456
Db     889 KLEEESAAKSKKVSEESVKFGEEKKTKAGEKTVQVESEPTSKKTIDTKDVGATEPADETP 948

Qy     457 KK--------KRTGSVHHTVAHPAGNLSNKK--------------VTPTASTQHDDEND 493
Db     949 KKKIIKKKTEKSDSSISQKSATDSEKVSKQKEQDEPTKPAVSETQMVTEADKSKKQKETD 1008

Qy     494 TENGLDTNM-----------HKTDVCQHVSEISTQRCSSKGKTAGLSKGKTHSAASTKYG 542
Db    1009 EKLKLDAEIAAKTKQEADEKSKLDAQEKIKKVSEDDAARKEKELN-DKLKLESEIATKKA 1067

Qy     543 G-------ESTRNGQNIHVLSA------EDQCQMETENSVLSHSAKVSPAEHDIQIMSDL 589
Db    1068 SADKLKLEEQAQAKKAAEVEAAKKQKEKDEQLKLDTEAASKKAAAEKLELEKQAQIKKAA 1127

Qy     590 HEQSLPKKK---KKQKLEVTRE---------------KQTMIDDIPMDIVELLAKNQHE 630
Db    1128 GADAVKKQKELDEKNKLEANKKSAAGKLKIEEESAAKSKQTVEEQAKLD-AQTKAKTA-E 1185

Qy     631 RQLMTETDCSDINRIQSKTTADD-----------------------DCVIVAAKDGSD 665
```

```
Db   1186 KQTKLEKDEKSTKESESKETVDEKPKKKVLKKKTEKSDSSISQKSETSKTVVESAGPSES 1245

Qy    666 YASSVFDTNSQQK----------------------------------SLASQSTQKELQG 691

Db   1246 ETQKVADAARKQKETDEKQKLEAEITAKKSADEKSKLEAESKLKKAAEVEAAKKQKEKDE 1305

Qy    692 HLALTTQESPHPQNFQSTQ-EQQTHLRMEEMVTIAASSPLFSHHDDQYIAEAPTEHWGRK 750

Db   1306 QLKLDTEAASKKAAAEKLELEKQSHIKKAAEVD-AVKKQKELEEKQRLESEAATK---KA 1361

Qy    751 DAKKLTWEQFKATTRNSPAATCG-----------AQFRPGIQAVDLTSTHVMGSSSNYAS 799

Db   1362 DAEKLKLEEQK-----KKAAEIALIEIQKEQEKLAQEQSRLEDEAKKSAEKQKLESETKS 1416

Qy    800 RQPVIAPLDRYAERAVNQV-------------HARNFPSTI---ATMEAS-KLCDRRNA 841

Db   1417 KQTEEAPKESVDEKPKKKVLKKKTEKSDSSISQKSKSAKSTVDAAETLESDFNLVEKKTV 1476

Qy    842 GQVVLYPKESMPAT 855

Db   1477 QKVEQSPDESTSAT 1490
```

RESULT 12
Q9N435
ID Q9N435 PRELIMINARY; PRT; 2083 AA.
AC Q9N435;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 231.2 KDA PROTEIN.
GN Y38B5A.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Sun H., Geisel C.;
RT "The sequence of C. elegans cosmid Y38B5A.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AC024760; AAF59460.2; -.
DR InterPro; IPR003962; FnIII_repeat.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR003599; Ig.
DR Pfam; PF00041; fn3; 2.
DR PRINTS; PR00014; FNTYPEIII.
DR SMART; SM00060; FN3; 3.
DR SMART; SM00409; IG; 3.
KW Hypothetical protein; Repeat.
SQ SEQUENCE 2083 AA; 231174 MW; 9270303ADE2D7C2F CRC64;

Query Match 3.7%; Score 205.5; DB 5; Length 2083;
Best Local Similarity 18.6%; Pred. No. 0.0006;
Matches 175; Conservative 135; Mismatches 435; Indels 197; Gaps 30;

```
Qy      1 MEIVAVDQEGARVVGTNCMLARGGTGAVAPVLELTATP---------------------- 38

Db    467 IELVPNEFYAFRIVAVNGF----GEGAPSEIIEVNTLDYDQEESFDFAGEEELKLDDVQV 522

Qy     39 RQDAAAEAGVDEPAQHQCEHFSIRGYVALLQKKDPKFCSLSRIFHDQKKCDEHKASSS-P 97

Db    523 NNEVVTEITIEESEVTIEEH-------RKLKKKSKK---------SKKTTDEPELDSEIA 566

Qy     98 FSVAKFRRWDCSKCLDKLKTSDNGTAPRTLPAKQNGTSDGCSITFVRSTFVPASVGSQKV 157

Db    567 LEVS-------SDITSSLEITTESTIPDTAPESQETLNVEIAVT---ETTV------QKI 610

Qy    158 SPSTQSSQGKNADRSTLPKSVQEGNDSKCNAPSGKNGAAEANTDSPMKDLQGPAQNYDVA 217

Db    611 TNPSDESAKKDVNEDTAVSSIVKKDDKDVN----KKSLPESGL-TTKKEIQGKPEKKIMK 665

Qy    218 ANVSEDNTSVDVGALPEVPQITWHIEVNGADQPPSTPKLSEVVLKRNEDENGKTEETLVA 277

Db    666 KKTEKADSSISETS----ETLTKDLTQTKQSEPEPAKRTTETSVQ--DEVKRKTETTSKS 719

Qy    278 EQCNLTKDPNPMSGKERDQVAEQCNLTKDPKPVSGQKCE-QICNEPCEEVVLKRSSKSKR 336

Db    720 KQ---TTEEHPQPGGKSDSSISSTSDASEVKQVQQSESEAQKVTEKPETAKLESKSKMTE 776

Qy    337 KTDKKLMKKQQHSKKRTAQADVSDAKLCRRKPKKVRLLSEIINANQVEDSRSDEVHREN- 395

Db    777 DTTK------ESDNKETVDEKPKKKVLKKKTEKSDSTISETSETSAVESAGPSESETQNV 830

Qy    396 AADPCEDDRSTIPVPMEVSMDIPVSNHT--------------VGEDGLKSSKNKT---- 436

Db    831 AAVDKEKKQKETDEKQKLEAEIAGKKSTEQKSKLEAEAKLKRAAEEDAAKKQKEKTEAAS 890

Qy    437 ----------------KRKYSDVVDDGSSL--MNWLNGKKK----------RTGSVHHT 467

Db    891 KKAAAEKLELEKQAQINKAAEADAVKKQNELDEQNKLEATKKLAAEKLKLEEQSAAKSKQ 950

Qy    468 VAHPAGNLSNKKVTPTASTQHDDENDTENGLDTNMHKT-------DVCQHVSEISTQRCS 520

Db    951 AAEEQAKLDAQTKAKAAEKQTGLEKDEKSNKDSGSNETVEEKPKKKVLKKKTEKSDSSIS 1010

Qy    521 SKGKT-------AGLSKGKTHSAASTKYGGESTRNGQNIHV-----LSAEDQCQMETENS 568

Db   1011 QKSDTSKTVAESAGSSESETQKVADATSKQKETDKKQKLEAEITAKKSADEKSKLETESK 1070

Qy    569 VLSHSAKVSPAEHDIQIMSDLHEQSLPKKKKKQKLEVTREKQTMID-----DIPMDIVEL 623

Db   1071 LIKAAEDAAKKQKEKEDKLKLEADVASKKAAAEKLEL--EKQAQIKKAAEADAVKKQKEL 1128

Qy    624 LAKNQHERQLMTETDCSDINRIQSKTTADDDCVIVAAK-----DGSDYASSVFDTNSQQK 678

Db   1129 AEKQKLESEAATKKAAAEKLKLEEQAQINKAAEADAVKKQKELDEKNKLEANKKSAAEKL 1188

Qy    679 SLASQSTQKELQGHLALTTQESPHPQNFQSTQEQQTHLRMEEMVTIAASSPLFSHHDDQY 738

Db   1189 KLEEESAAKSKQ---TVEEQAKLDAQTKEKTAEKQTGLEKDDKST----------KDSE 1234

Qy    739 IAEAPTEHWGRKDAKKLTWEQFKATTRNSPAATCGAQFRPGIQAVDLTSTHVMGSSSNYA 798

Db   1235 SKETVDEKPKKKVLKKKTEKSDSSISQKS----------------VTSKTVVESGG--- 1274

Qy    799 SRQPVIAPLDRYAERAVNQVHARNFPSTIATMEASKLCDRRN 840

Db   1275 ---PSESETQKVADAARKQKETDEKQKLEAEITAKKSADEKS 1313
```

RESULT 13
O76891
ID O76891 PRELIMINARY; PRT; 5327 AA.
AC O76891;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE EG:49E4.1 PROTEIN.
GN FUTSCH OR EG:49E4.1 OR CG3064.

```
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Papagiannakis G., Spanos L., Siden-Kiamos I., Louis C.;
RT   "Sequencing the distal X chromosome of Drosophila melanogaster.";
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Benos P.;
RL   Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL031128; CAA20006.1; -.
DR   FlyBase; FBgn0015390; futsch.
SQ   SEQUENCE   5327 AA;  575942 MW;  FEFEE23A118FF33A CRC64;


  Query Match             3.7%;  Score 204.5;  DB 5;  Length 5327;
  Best Local Similarity   19.1%;  Pred. No. 0.0025;
  Matches  225;  Conservative  174;  Mismatches  457;  Indels  319;  Gaps   51;

Qy      31 VLELTATPRQDAAAEAGVDEPAQHQCE---HFSIRGYVA----LLQKKDPKFCSLSRIFH 83
           | | :  | ::|:  | | |  : : |     | |  ||    |  |:  :  |::
Db    1967 VAEKSPLPSKEASRPASVAESIKDEAEKSKEESRRESVAEKSPLPSKEASRPASVAESIK 2026

Qy      84 DQ-KKCDEH--------------KASSSPFSVAKFRRWDCSKCLDKLKTSDNGTAPRTLP 128
           |: :|  |               | :| | |||:  : :  |  :: :  ::       ||
Db    2027 DEAEKSKEESRRESVAEKSPLPSKEASRPASVAESIKDEAEKSKEESR-RESAAEKSPLP 2085

Qy     129 AKQNGTSDGCSITFVRSTFVPASVG---SQKVSPSTQSSQGKNADRSTLPKSVQEGNDSK 185
           :|:                 ||||      :   | : |: ::   |   :|: :|: |
Db    2086 SKEASR--------------PASVAESVKDEADKSKEESRRESMAESGKAQSI-KGDQSP 2130

Qy     186 CNAPSGKNGAAEANTDSPMKDLQGPAQNYDVAANVSEDNTSVDVGALPEVPQITWHIEVN 245
           |     ||:  | |:|  : |::   || :|: |:   |   |                |
Db    2131 LKEVSRPESVAESVKDDPVKSKE-PSRRESVAGSVTADSARDDQSPL----------ESK 2179

Qy     246 GADQPPS-TPKLSEVVLKRNEDENGKTEETLVAEQCNLTKDPNPMSGKERDQVAEQCNLT 304
           || :| |    : :   |:      |||  :          |      |   :| :  ::|
Db    2180 GASRPESVVDSVKDEAEKQESRRESKTESVI---------PPKAKDDKSPKEVLQPVSMT 2230

Qy     305 KDPKPVSGQKCEQICNEPCEEVVLKRSSKSKRKTDKKLMKKQQHSKKRTAQADVSDAKLC 364
           :  :  : |  :    |   | : :   |  : :|  :  :: |:  :    :   |
Db    2231 ETIREDADQPMKPSQAESRRESIAESIKASSPRDEKSPLASKEASRPGSVAESI---KYD 2287

Qy     365 RRKPKKVRLLSEIINANQVEDSRSDEVHRENAADPCEDDRSTIPVPMEVSMDIPV-SNHT 423
           ||: ::          :|  ::   ||:      :|:| :     ||  : | |:|
Db    2288 LDKPQIIK----------DDKSTEHSRRESL-----EDKSAVTSEKSVSRPLSVASDHE 2331

Qy     424 VG---EDGLKSS---KNKTKRKY-----------------SDVVDDGS---SLMNWLNG 456
             ||  |||   |:|::  :                   :||:  |   ||     |
Db    2332 AAVAIEDDAKSSISPKDKSRPGFVAETVSSPIEEATMEFSKIEVVEKSSLALSLQGGSGG 2391

Qy     457 KKK--------RTGSVHHTVAHPAGNLSNKKVTPTAS-----TQHDDENDTENGLDTNMH 503
           | :          |   | ||       |   |||| :      |        : ||  :
Db    2392 KLQTDSSPVDVAEGDFSHAVA------SVSTVTPTLTKPAELAQIGAAKTVSSPLDEALR 2445

Qy     504 KTDVCQHVS------EISTQRCSSKGKTAGLSKGKTH---------SAASTKYGGESTRN 548
           :|:|      | :::  :|: |:  : |  :              ||  |   |  |:
Db    2446 TPSAPEHISRADSPAECASEEIASQDKSPQVLKESSRPAWVAESKDDAAQLKSSVEDLRS 2505

Qy     549 GQNIHVLSAE---DQCQMETENSVLSHS----AKVSPAEH-----DIQIMSDLHEQSLP- 595
           | | |        || :| :  :    |:   ||       |:: :|   | |
Db    2506 ----PVASTEISRPASAGETASSPIEEAPKDFAEFEQAEKAVLPLTIELKGNLPTLSSPV 2561

Qy     596 -------KKKKKQKLEVTREKQTMIDDIPMDIVELLAKNQHERQLMTETDCS-------D 641
                 :  :  |::: :   : ||: |  ::   |: : |    :  | :       |
Db    2562 DVAHASVQPAELSKVDIEKTASSPIDEAPKSLIGSPAEERPESPAESAKDAAESVEKSKD 2621

Qy     642 INR---IQSKTTAD-----------------DD------------------------- 654
           :|   :   | ||                  ||

Db    2622 ASRPPSVVESTKADSTKGDISPSPESVLEGPKDDVEKSKESSRPPSVSASITGDSTKDVS 2681

Qy     655 ---CVIVAAKDGSDYA----------SSVFDT--NSQQKSLASQSTQKELQGHLALTTQE 699
             |: : ||  | |           || |    |  || :|| :||: :  || : :
Db    2682 RPASVVESVKDEHDKAESRRESIAKVESVIDEAGKSDSKS-SSQDSQKDEKSTLA-SKEA 2739

Qy     700 SPHPQNFQSTQEQQTHLRMEEMVTIAASSPLFSHHDDQYIAEAPTEHWGRKDAKKLTWEQ 759
           |      :|:::            ||:  |:            | |     |:|  :
Db    2740 SRRESVVESSKDDAEKSESRPESVIASGEPV-----------PRESKSPLDSKDTS--- 2784

Qy     760 FKATTRNSPAATCGAQFRPGIQAVDLTSTHVMGSSSNYASRQPVIAPLDRYAERAVNQVH 819
                        |||      :|:       |   ||:  :|   :   :   :
Db    2785 ----------------RPGSVVESVTAE---DEKSEQQSRRESVAESVKADTKKDGKSQ 2824

Qy     820 ARNFPSTIATMEASKLCDRRNAGQVVLYPKESMPATH--LLRMMDPSTLASFPNYGTSSR 877
             : ||::   |  |  | :   :     ::|:  :|  :  | | | :         ||
Db    2825 EASRPSSVD--ELLKDDDEKQESR-----RQSITGSHKAMSTMGDESPMDKADKSKEPSR 2877

Qy     878 NQ--MESQLHNSQYAHNQYKGSTSTSYGSNLNG------KIPLTFEDLSRHQLHDLHRPL 929
             :  ||  | :        ||   |   ::       | ||  :::|| :
Db    2878 PESVAESIKHENTKDEESPLGSRRDSVAESIKSDITKGEKSPLPSKEVSRPE-------- 2929

Qy     930 RPHPRVGVLGSLLQKEIANWSENCGTQSGYKLGVSTGITSHQMNRKEHFEALNSGMFSAK 989
                 |:||:  ::  :  |:       |   |   ||   : |||    :
Db    2930 ------SVVGSIKDEKAESRRES--VAESVKPESSKDATSAPPS-KEHSRPES------- 2973

Qy     990 WNALQLGSVSSSADFLSARN-SIAQSWTRGKGKMV 1023
                 |||:    |  ::|  |:| |    | :|
Db    2974 ----VLGSLKDEGDKTTSRRVSVADSIKDEKSLLV 3004


RESULT  14
Q9VC00
ID   Q9VC00      PRELIMINARY;      PRT;  2768 AA.
AC   Q9VC00;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CG13648 PROTEIN.
GN   CG13648.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
```

```
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003750; AAF56376.1; -.
DR   FlyBase; FBgn0039257; CG13648.
DR   InterPro; IPR001007; VWFC.
DR   Pfam; PF00093; vwc; 1.
DR   SMART; SM00214; VWC; 4.
DR   PROSITE; PS01208; VWFC; 1.
SQ   SEQUENCE   2768 AA;  294032 MW;  CA929A21774E4684 CRC64;

  Query Match             3.7%;  Score 204;  DB 5;  Length 2768;
  Best Local Similarity   19.7%;  Pred. No. 0.0011;
  Matches  181;  Conservative  118;  Mismatches  345;  Indels  274;  Gaps   40;

Qy      29 APVLELTATPRQDAAAEAGVD-------------------EPAQHQCEHFSIRGYVALL 68
           |||   || | :|   :  |:                   ::::  |  | |
Db    1550 APVAVTTAAPSKDDGEQKPVEVEEKPIEDGQKPIEDETSTPTSSENEIEPESDRATTIAP 1609

Qy      69 QKKDPKFCSLSRIFHDQKKCDEHKASSSPFSVAKFRRWDCSKCLDKLKTSDNGTAPRTL- 127
           |::|   |      |:       :: :| |             |: | :    | |:
Db    1610 SKEEPSEPSTGAPTKDE---PAEPSTDAPES-----------DESKETPESEVPTTVA 1653

Qy     128 PAKQNGTSDGCSITFVRSTFVPASVGSQKVSPSTQSSQGKNADRSTLPKSVQEGNDSKCN 187
           || :                :| |  :    |:  |:     |       :   :  :
Db    1654 PAGEK--------------IPTSSITPDEEPTATSAPVAKPDEDV---EKETSTEIPTD 1695

Qy     188 APSGKNGAAEANTDS-----PMKDLQGPAQ---NYDVAANVSEDNTSVDV--------GA 231
           ||:       ::||     | |  : |||     |:    :   || :|
Db    1696 APASSEEDENSSTDQIPSEVPEKKPETPAQTPEEGDIVGATAAPTTSDEVPPVQRLPEEV 1755

Qy     232 LPEVPQITWHIEVNGADQPPSTPKL------------SEVVLKRNEDENGKTEETLVAE 278
           | |:|| :    :   |:  : | :                :  |   :| :   |||    |
Db    1756 LAEIPQPSTETGIKQQDETTAAPSIDRKEPYVTEIDEEATTVAPISEKDEKPTEEEKPVE 1815

Qy     279 QCNLTKDPNPMSGKERDQVAEQCNLTKDPKPVSGQKCEQICNEPCEEVVLKRSSKSKRKT 338
           |    ::|:    :|:::  ||   |:  ||| :  |    |  |||          | |
Db    1816 QKPTGEEPS--EEEEKEKPIEQDVSTEG--PVSTEASEAGSTESSEEV--------KPST 1863

Qy     339 DKKLMKKQQHSK-KRTAQADVS-----DAKLCRR---KPKKVRLLSEIINANQVEDSRSD 389
           : :: :| :  :   |||| |        :|  :   ||     :   : :  | :
Db    1864 EGEVAEKPEDKQPSSTAQAPVETIPEISTELPAQDGDKPTSEAPVDSDEDTSAPSDEKIP 1923

Qy     390 EVHRENAADP---------CEDDRSTIPVPMEVSM--DIPVSNHTVGEDGLKSSKNKTKR 438
           | |   |           |:|    |   | |    :| : :   ||  |:  ::|
Db    1924 SVSGEEVEGPEVTTASPQAAEEDELKTPAESEPSSTDKVPETEYQKPEDETKA--DETPE 1981

Qy     439 KYSDVVDDGSSLMNWLNGKKKRTGSVHH----TVAHPAGNLSNKKVTPTASTQHD----- 489
             : | |  :|    : |    | :      | | |      ::: || :   :
Db    1982 SVTQVSDVATSTSAPVAG-----GDIEKDEQATTASPE---EEEEIKPTIAPAAEIPQPS 2033

Qy     490 -----DENDTENGLDTNMHKTDVCQHVSEISTQRCSSKGKTAG-LSKGKTHSAASTKYGG 543
                || : |:|      ::|  | : ||:          |:| : :  |  || ||
Db    2034 EKEPVDEQEVESGTKATPAESD-GQPIDEIAP-------ATSGPIDEAST--AAPTK--- 2080

Qy     544 ESTRNGQNIHVLSAEDQCQMETENSVLSHSAKVSPAEHDIQI--------MSDLHEQSLP 595
                                 | |    ||  ||| || :|        ::|  | : |
Db    2081 ---------------------EESTTVASA-ASPAVHDDEIKDVTTTQPVADEKEVAAP 2117

Qy     596 KKKKKQKLEVTREKQTMIDDIPMDIVELLAKNQHERQLMTETDCSDINRIQSKTTADDDC 655
           : : |  ::|: :  |  ||              |:|  ||   :       |  |||
```

```
Db    2118 QDETKTSIDVSTDSPTAQDD-------------EKQDKTEAPVAPTT--VSSPTADSAA 2161

Qy     656 -----------VIVAAKDGSDYASSVFDTNSQQKSLASQSTQKELQGHLALTTQE---- 699
                       | :  |   |  |     ::     || ||: | |  : :: |:
Db    2162 DSSTPTVEVPSPVEIDTKPMDDIMSQTIAPHTAD-GAASTSTEDEDQAPVTVSPQDAEKT 2220

Qy     700 --SPHPQNFQSTQEQQTHLRMEEMVTIAASSPLFSHHDDQYIAEAPTEHWGRKDAKKLTW 757
             || ||:   |  :      :|:  | ::||    ||
Db    2221 PVSPAPQDSDKTPSSEAPQDADEIP--ATATPL----DD---------------------- 2253

Qy     758 EQFKATTRNSPAATCGAQFRPGIQAVDLTSTHVMGSSSNYASRQPVIAPLDRYAERAVNQ 817
                  |   ||   |   |: |          :  :    |  ||||
Db    2254 --------NKIPATVAPQTDDGVPAT--------AAPLDEDKIQTTAAPLDE-------- 2289

Qy     818 VHARNFPSTIATMEASKL 835
                |||  |  ::  |:
Db    2290 ---EKIPSTAAPLDDEKI 2304


RESULT  15
Q9W3Z0
ID   Q9W3Z0      PRELIMINARY;      PRT;  2951 AA.
AC   Q9W3Z0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   CG3950 PROTEIN.
GN   CG3950.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
```

```
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003438; AAF46171.1; -.
DR   FlyBase; FBgn0029875; CG3950.
SQ   SEQUENCE   2951 AA;  332269 MW;  85CCC0ABBDAC8A3B CRC64;


  Query Match             3.7%;  Score 204;  DB 5;  Length 2951;
  Best Local Similarity   20.1%;  Pred. No. 0.0012;
  Matches  213;  Conservative  161;  Mismatches  463;  Indels  222;  Gaps   48;

Qy      78 LSRIFHDQKKCDEHKASSSPFSVAKFRRWDCS-------KCL-------DK-LKTSDNGT 122
           :||  |  :|  | |  |:|||    | :| :        ||        || |:|::|
Db    1838 VSRFTHSAEKVKEPKV-SAPFS----REFDVNAKIPENDDCLLSINQKVDKFLRTAENVI 1892

Qy     123 APRTLPAK------------QNGTSDGCSITFVRST--FVPASVGSQKVSPSTQSSQGK- 167
            | :||::             :    | |:::  :    |:     ::|::|:      :
Db    1893 RPTSLPSRPEIERPGLEEIDEELLRDDCTLSVSQKVHKFIDT---AEKLAPTMPQKSPRL 1949

Qy     168 --NADRSTLPKSVQEGNDSKCNAPSGKNGAAEANTDSPMKDLQGPAQ---NYDVAANVSE 222
             | :|    :|  |    : : |       | : |: ::|    :|     ::   |::
Db    1950 VANIERHISRQSEPERELDEESEP-------ELDRDTDVEDDDQTSQLETEEEITQTVTK 2002

Qy     223 DNTSVDVGALPEVPQITWHIEVNGADQPPSTPKLSEVVLKRNEDENGKTEETLVAE-QCN 281
           |       | |  | |         |      ||     |::     | |   | : |
Db    2003 KET------LKEFKQQTKETRETRRDSKAEPEKLQ----KKSPQTKVKEESARVPKYQAK 2052

Qy     282 LTKDPNPMSGKERDQVAEQCNLTKDPKPVSGQKCEQICNEPCEEVVLKRSSK-SKRKTDK 340
           :::  :    |:: |   :    : |     |   :: :||  | | ||  |  ::::
Db    2053 VSQKVSQWEPKKQPQREPKVTQKETPLEPKKQPLSKVKDEP--EKVNKREPKVPQKESQT 2110

Qy     341 KLMKKQQHSKKRTAQADVSDAKLCRRKPKKVRLLSEIINANQVEDSRSDEVHRENAADPC 400
           || :: :   |:| |           ::|:|  |         | ||       |   :|
Db    2111 KLKEEPERVTKKTPQ----------KEPRKEPL-------RQSEDEPEFSPEEEFDDEPL 2153

Qy     401 EDDRS-TIPVPMEVSMDIPVSNHTVGE--DGLKSSKNKTKRKYSDVVDDGSSLMNWLNGK 457
             :: |  : |:   ||       |:     |||   :  | :       |   | :  :
Db    2154 PMTKTHTTAIEMKRQKDILNRPSVFGQRTPERKSSTTPSPTKLNGTRGRPSPSTNLITEE 2213

Qy     458 KKRTGSVHHTVAHPAGNLSNKKVTPTASTQHDDENDTENGLDTNMHK-----------T 505
           |:   :    |: |      :| ||:|::           :   | :
Db    2214 KRSYRNQVTNVSKP----GTRKTTPSANSPAQSPPPKTTSISKRMEQISQQSWVVQDVDV 2269

Qy     506 DV-------CQHVSEISTQRCSSKGKTAGLSKGKTHS---AASTKYGGESTRNGQNIHVL 555
           ||          |:||    :  |   :  ||:  : |   : :| |    | |    :|
Db    2270 DVEVVGPAPPSHISEKPQGKSPSPTSSRSLSRSPSRSPKKSITTTY----TTNTTGRNVA 2325

Qy     556 SAEDQCQMETENSVLSHSAKVSPAEHDIQIMSDLHEQSLPKKKKKQKLEVTR----EKQT 611
           |  :  :   |  | |      |:  |       | :|  :  ::  ||: :     | :
Db    2326 SRRNVFEPVHETHVDSEPTGRRPSYMD-------HTKSSLEHIRRDSLEINKSHYSRKSS 2378

Qy     612 MIDDIPMDIVELLAKNQHERQLMTETDCSDINRIQSKTTADD--DCVIVAAKDGSDY--- 666
           | || |::      :| :           |: |  |   ||:     :    : ||
Db    2379 MEDDSPVE-----PRNPNSSVKF------DVPRKSSSRGADEPRKTSLKGKDEDSDLELE 2427

Qy     667 ASSVFDTNSQQK---SLASQSTQKELQGHLALTTQESPHPQNFQSTQEQQTHLRMEEMVT 723
             :||    :|   ::||   :: ::  : |          :|  :     |   :   |
Db    2428 IEEIFDLQRLEKLLETVASYEMRRRIRAQMRLIR------KNMINAGTTTTITTITTSTT 2481

Qy     724 IAASSPLFSHHDDQYIAEAPTEHWGRKDAKKLTWEQFKATTRNSPAATCGAQFRPGIQAV 783
             ||||     ||  | |                   :  |:     |   | :  :: |
Db    2482 PGKSSPLPKIRRDQSPAGAA----------------EVKTKEVRTTTSRRQQQQRVEQV 2524

Qy     784 DLTSTHVMGSSSNYASRQPVIAPLDRYAERAVNQVHARNFP---------STIATMEASK 834
           | |:    | :| :   :| : | :| |  |  :   |  |          ||  | : :
Db    2525 DSTTPIAPGKTSPHG--KPPVKPRERSASPAQKR---RISPPGKQSPGDRSTTTTTKVTT 2579

Qy     835 LCDRRNAGQVVLYPKESMPA--------THLLR--MMDPSTLASFPNYGTSSRNQMESQL 884
             | |         | ||        : :|:     | |  | |  |::|     |
Db    2580 TSTTRGA--------PSKPAQGPIWADRSKVLKGHATVPQTNGSTPRKGSTSSTTSSS-- 2629

Qy     885 HNSQYAHNQYKGSTSTSYGSNLNGKIPLTFEDLSRHQLHDLHRPLRPHPRVGV----LGS 940
             :        ||:||  |  | :     ||            : |    |:    : :

Db    2630 --GKITRTMTSSSTTTSSSSTTNTRNKQREEDSITSSY-----GVGPTDENGLPLFGIRA 2682

Qy     941 LLQKEIANWSENCGTQ---SGYKLG---VSTGITSHQMNRKEHFEALNSGMFSAKWNALQ 994
           | :|      | |  |:   :|| :     |   :  :  |||   : |:   :|    ||
Db    2683 LKKKATPPAEEPCETKQEVTGYVIEEQFYSDNKSPPRHERKELIYSSNADELAAIKQQLQ 2742

Qy     995 LGSVSSSADFLSARNSIAQSWTRGKGKMVHPLD-RFVRQ 1032
                ||   | ||  : : : : : :   | | |:||:
Db    2743 -DEDDSSPPLLDAR--VVREFKKVESQQSLPEDARYVRR 2778


Search completed: September 16, 2002, 22:34:15
Job time: 654 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: September 16, 2002, 20:09:51 ; Search time 65.82 Seconds
(without alignments)
1783.728 Million cell updates/sec

Title: US-09-828-068-2
Perfect score: 5526
Sequence: 1 MEIVAVDQEGARVVGTNCML.....NKNPADFTTISNDNEYMDYR 1057

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	393	7.1	1096	21	AAY84833	Amino acid sequenc
2	221	4.0	1982	22	ABG16404	Novel human diagno
3	221	4.0	1982	22	ABG19656	Novel human diagno
4	220	4.0	1325	22	ABB29725	Peptide #2376 enco
5	220	4.0	1325	22	ABB34897	Peptide #2403 enco
6	220	4.0	1325	22	ABB20314	Protein #2313 enco
7	220	4.0	1325	22	AAM55707	Human brain expres
8	220	4.0	1325	22	AAM68085	Human bone marrow
9	220	4.0	1325	22	AAM03645	Peptide #2327 enco
10	217	3.9	1165	22	AAG70791	S cerevisiae apopt
11	208	3.8	1528	18	AAW20056	C. elegans UNC-53
12	208	3.8	1583	18	AAW20057	C. elegans UNC-53
13	206.5	3.7	2070	22	ABG03703	Novel human diagno
14	204	3.7	2768	22	ABB68397	Drosophila melanog
15	204	3.7	2951	22	ABB60291	Drosophila melanog
16	201.5	3.6	647	10	AAP93971	Human carbastatin-
17	201.5	3.6	665	21	AAB18278	Plasmodium falcipa
18	201.5	3.6	1784	22	ABB71460	Drosophila melanog
19	201	3.6	2344	22	AAU37120	Staphylococcus aur
20	200	3.6	3111	22	ABB60327	Drosophila melanog
21	198.5	3.6	1690	22	ABB61144	Drosophila melanog
22	198.5	3.6	1690	22	ABB61173	Drosophila melanog
23	198	3.6	1552	21	AAY85577	C. elegans UNC-53
24	197.5	3.6	754	21	AAY70008	Human Protease and
25	197	3.6	694	17	AAR85596	Human calpastatin.
26	196.5	3.6	2515	22	ABB12281	Human secreted pro
27	196.5	3.6	2515	22	AAM80268	Human protein SEQ
28	196.5	3.6	2515	22	AAM80269	Human protein SEQ
29	196.5	3.6	2519	22	ABG16636	Novel human diagno
30	195	3.5	1730	22	ABB68370	Drosophila melanog
31	194.5	3.5	673	11	AAR05701	Human carbastatin
32	193.5	3.5	1803	22	ABB65391	Drosophila melanog
33	192.5	3.5	2025	22	AAU34207	Staphylococcus aur
34	192.5	3.5	3158	22	AAU37018	Staphylococcus aur
35	192	3.5	3201	22	ABB62899	Drosophila melanog
36	191.5	3.5	982	12	AAR13320	Murine Natural Kil
37	191.5	3.5	2633	22	ABG06505	Novel human diagno
38	188.5	3.4	2897	22	ABB58514	Drosophila melanog
39	188	3.4	2505	22	AAM40002	Human polypeptide
40	187.5	3.4	2586	22	ABB66878	Drosophila melanog
41	187	3.4	1420	22	ABB63410	Drosophila melanog
42	187	3.4	1489	22	ABB59948	Drosophila melanog
43	187	3.4	1863	17	AAR81505	BRCA1 mutant from
44	187	3.4	3257	22	ABB67502	Drosophila melanog
45	184.5	3.3	2663	22	AAM39097	Human polypeptide

ALIGNMENTS

RESULT 1
AAY84833
ID AAY84833 standard; Protein; 1096 AA.
XX
AC AAY84833;
XX
DT 08-AUG-2000 (first entry)
XX
DE Amino acid sequence of the EMF-1 gene.
XX
KW Embryonic flower gene-1; EMF-1; reproductive development; flowering;
KW early flowering; uniform flowering.
XX
OS Arabidopsis thaliana.
XX
PN WO200022132-A2.
XX
PD 20-APR-2000.
XX
PF 08-OCT-1999; 99WO-US23543.
XX
PR 09-OCT-1998; 98US-0169696.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Sung ZR, Aubert D, Chen L;
XX
DR WPI; 2000-317987/27.
DR N-PSDB; AAA14873.
XX
PT An isolated nucleic acid molecule comprising embryonic flower genes is
PT useful for modulating reproductive development in plants -
XX
PS Claim 7; Page 34-38; 38pp; English.

```
XX
CC   The present sequence is encoded by the Arabidopsis thaliana embryonic
CC   flower gene-1 (EMF-1). The EMF genes play an important role
CC   in reproductive development in plants. Control of the expression of
CC   the genes is therefore useful in controlling flowering and other
CC   functions in plants. The EMF-1 polynucleotides are useful for modulating
CC   reproductive development in plants. For example, enhanced expression
CC   of EMF-1 is useful to prevent flowering in grasses or vegetable crops.
CC   Controlling or inhibiting expression of the genes is useful to promote
CC   early flowering in forest trees and agronomical crops or to ensure
CC   uniform flowering in various agronomic crops or ornamentals.
XX
SQ   Sequence   1096 AA;


  Query Match             7.1%;  Score 393;  DB 21;  Length 1096;
  Best Local Similarity   21.0%;  Pred. No. 2e-21;
  Matches  250;  Conservative  189;  Mismatches  403;  Indels  348;  Gaps   58;

Qy      55 QCEHFSIRGYVALLQKKDPKFC------SLSRIFHDQKKCDEHKASSSPFSVAKFRRWDC 108
           :|:|||:||:||  :::| : |      |:| :       |:   :    || ||| | |
Db      25 kcdhfsmrgfvaetrerdlrkcwpfseesvslv-------dqqsytlptlsvpkfrwwhc 77

Qy     109 SKCLDKLKTSDNGTAPRTLPAKQNGTSDG-CSITFVRSTFVPASVGSQKVSPSTQSSQGK 167
             |: :   :   |:      | : |  |:   :| |   ::   :    |
Db      78 mscikdi----dahgpkdcglhsnskaignssvieskskfnsltiidhekekktdi---- 129

Qy     168 NADRSTLPK-SVQEGNDSKC---------NAPSGKNGAAEANTDSPMKDLQGPAQNYDVA 217
           || :   |  |   || :           | |      :|  |  : |  | |   |
Db     130 -adnaieekvgvncenddqtattflkkargrpmg-----asnvrsksrklvspeq---vg 180

Qy     218 ANVSED---NTSVDVGALPEVPQITWHIEVNGA--------DQPPSTPKLSEVVLKRNED 266
           | |::     |:|: :  |   :   :   |:        | ||   |  : :    |
Db     181 nnrskeklnkpsmdisswkekqnvdqavttfgsseiagvvedtppkatknhkgirglmec 240

Qy     267 ENGKTEETLVAEQCNLTKDPNPMSGKER---DQVAEQCNLTKDPKPVSGQKCEQICNEPC 323
           :|| :|   :|          ||| :|    :|     |  : |   |   :
Db     241 dngssesinla----------msglqrrksrkvrllsellgntktsggsnirk------ 283

Qy     324 EEVVLKRSSKSKRKTDKKLMKKQQH-SKKRTAQADVSD--AKLC---------------- 364
           ||  ||: |   ||  :||: :  : |:  :     |:  :| |
Db     284 eesalkkesvrgrk--rkllpennyvsrilstmgatsenaskscdsdqgnsestdsgfdr 341

Qy     365 -----RRKPKKVRLLSEIINANQVEDSRSDEVHRENAADPCEDDRST-----------IP 408
                ::: :: ::: | : :   | |:  |  :|: ||| :  |||           :|
Db     342 tpfkgkqrnrrfqvvdefvpslpcetsq--egikehdadpsk--rstpahslftgndsvp 397

Qy     409 VPMEVSMDIPVSNHTVGEDGLKSSKNKTKRKYSDVVDDGSSLM----NWLNGKK--KRTG 462
           |       | :  |  |   |  | |||:    |:|:| | :    | ::| :    ||
Db     398 cp-------pgtqrt--erklslpkkktkk---pvidngkstvisfsngidgsqvnshtg 445

Qy     463 SVHHTVAHPAGNLSNKKVTPTASTQHDDENDTENGLDTNMHKTD--VCQHVSEISTQRCS 520
              :||:     |: |:|                ||  |   :|    :::|::: :  :
Db     446 psmntvsqtrdllngkrv--------------gglfdnrlasdgyfrkylsqvndkpit 490

Qy     521 S---------KGKTAGLSKGKTHSAASTKYGGESTRNGQNIHVLSAEDQCQMETENSVLS 571
           |         : : |  :  :  |::|   |   | |  : ::   :       :|  :
Db     491 slhlqdndyvrsrdaepnclrdfsssskssggwlrtg--vdivdfrnnnhntnrssfsn 548

Qy     572 HSAKVSPAEHDIQIMSDL-----------------HEQSLPKKKKKQKLEVTREKQTMI 613
              :  |:  ::  :| :                        |: :   : | | |:|
Db     549 lklryppsstevadlsrvlqkdasgadrkgktvmvqehhgaprsqshdrkettteeqnn- 607

Qy     614 DDIPMDIVELLAKNQHERQLM-TETDCSDINRIQSKTTADDDCVIVAAKDGSDYASSVFD 672
           |||||:||||:||||:|| |   | | |:    |       : :::   :  |   |: |
Db     608 ddipmeivelmaknqyerclpdkeedvsnkqpsqetahksknallidlnetydngisled 667

Qy     673 TNSQQ--KSLASQSTQKELQGHLALTTQESPH----------PQN---FQSTQEQQTHLR 717
           |: :  |  :| : ::|   |  :  |:: |          |     |  ||| :
Db     668 nntsrppkpcssnarree---hfpmgrqqnshdffpisqpyvpspfgifpptqenr---- 720

Qy     718 MEEMVTIAASSPLFSHHDDQYIAEAPTEHWGRKDAKKLTWEQFKATTRNSPAATCGA--- 774
                 |||  || |: |::   ||   | ::    ::   :|        || :

Db     721 --------assirfsghncqwlgnlptv--gnqnpspssfrvlra------cdtcqsvpn 764

Qy     775 QFRPGIQAVDLTSTHVMGSSSNYASRQPVIAPLDRYAERAVNQVHARNFPSTIATMEASK 834
           |:|         ::| :  ||       :| |  :|   ::|   : | | |::  :||
Db     765 qyr--------eashpiwpss-------mippqsqykpvslninqstn-pgtls--qasn 806

Qy     835 LCDRRNAGQVVLYPKE---------------------------------SMPATHLLRMM 861
             :  |   |    |:                                 |:|| ||| ::
Db     807 nentwnlnfvaangkqkcgpnpefsfgckhaagvssssrpidnfssessipalhllsll 866

Qy     862 DPSTLASFP--NYGTS--SRNQMESQLHNSQYAHNQYKGSTSTSYGSNLNGKIPLTFEDL 917
           ||   :: |   :| :  ::        : ::   |   |: ::| :    :|| |:
Db     867 dprlrsttpadqhgntkftkrhfppanqskefielqtgdssksaystk---qip--fdly 921

Qy     918 SRHQLHDLHRPLRP-HPRVGVLGSLLQKEIANWSENCGTQSGYKLGVSTGITSHQMNRKE 976
           |:    :  |   |  | :|   | |  : |:|| :                ||  : :
Db     922 skrftqepsrksfpitppigt--sslsfqnaswsph----------------hqekktk 962

Qy     977 HFEALNSGMFSAKWNALQLGSVSSSAD-----FLSARNSIAQSWTRGKGKMVHPL----- 1026
             :      |:  :|  :   :|| |      | | ||           |: ||
Db     963 rkdt-----fapvynthekpvfassndqakfqllgasns----------mmlplkfhmt 1006

Qy    1027 DRFVRQD---------------------ICITNKNPADFTTISNDNEYM 1054
           |:  :|                      :|  |:||||||     | ||
Db    1007 dkekkqkrkaescnnnasagpvknssgpivcsvnrnpadftipepgnvym 1056


RESULT   2
ABG16404
ID   ABG16404 standard; Protein; 1982 AA.
XX
AC   ABG16404;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #16395.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS80591.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 46763; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
```

```
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   1982 AA;


  Query Match             4.0%;  Score 221;  DB 22;  Length 1982;
  Best Local Similarity   18.6%;  Pred. No. 1.5e-07;
  Matches  210;  Conservative  171;  Mismatches  423;  Indels  328;  Gaps   53;

Qy      83 HDQ---KKCDEHKASSSPFSVAKFRRWDCSKCLDKLKTSDNGTAPRTLPAKQNGTSDGCS 139
           ||:    ||  |    :     | |      |::||  ::: : |          |||  |
Db     946 hdevvdvkchshspfrnergkgkirhh--iscieklsktesisvp---------tsdhrs 994

Qy     140 ITFVRSTFVPASVGSQKVS---------PSTQSS---QGKNADRSTLPKSVQEG------ 181
           :       : |:  : |||          |   ||    |: :    :  |:: |
Db     995 l-------ieanqsnskvseldtiyctlprksssflihgrqsgskimaaslrngpppfqi 1047

Qy     182 ------------NDSKCNAPSGKNGAAEANTDSPMKDLQGPAQNYDVAANVSEDNTSVD 228
                       |    ::|   |  ::  :||    |: |     :    |  :|||
Db    1048 knnvedamgnymlnkfspsspesanecskvlsdsa---leapeatermtnvkssgstsvr 1104

Qy     229 VGALP-------EVPQITWHIE----------VNGADQPPSTPKLSEVVLKRNEDENGKT 271
           | ||         |    |              :| |    ||:  | ::   | ::
Db    1105 kgplpflinramscpsgephastgregrkkpltsgmdaseltpraweriispvesdss-- 1162

Qy     272 EETLVAEQCNLTK------------------------------DPNP----MSGKERDQ 296
                  |:|||                                || |    :|||||  :
Db    1163 -----vrdcsltkrqhqkenfqeytekegkmaasrrsvfalsnedplpfcsdlsgkergk 1217

Qy     297 VAEQCNLTKDPKPVSGQKCEQICNEPCEEVVLKRSSKSKRKTDKKLMKKQQHSKK----- 351
           :   |     ||| :     |  | |||    :  :: : |     ||:::
Db    1218 tlhkvk-ttstfsvsgde----dnvkclevvsiyytlprkpskkfcnllqqytqntnlli 1272

Qy     352 RTAQADVSDAKLCRRKPKK------------VRLLSEIINANQVEDSRSDEVHRENAADP 399
           : | :          | |:               |   :|::|   : :    | :   |
Db    1273 espqvetetfpnalekdkqnystreqsgtpscenlkmsvnsdqtlttenmtafrlsnrgp 1332

Qy     400 CEDD-------RSTIPVPMEVSMDIPVSNHTVGEDGLKSSKNKTKRKYSDVVDDGSSLMN 452
                     : : :| | |    : :  : :  |  |:|| :|
Db    1333 laptlqemasveaavslpeeeskareifsdnlaktplgdsenkker-------------- 1378

Qy     453 WLNGKKKRTGSVHHTVAHPAGNLSNKKVTPTASTQHDDENDTENGLDTNMHKTDVCQHVS 512
              ||| :: ::| ::     |:| :| :       :  |   : |   : : ::    :
Db    1379 ---gkklqsetlhtslmlqrknvseek-sencqqsinssnsgpsslpa-lsevnignsqt 1433

Qy     513 EISTQRCSSKGKT-----AGLSKGKTHSAASTKYG--GESTRNGQNIHVLSAEDQCQMET 565
             |:  |:  |:      :|    | |:: :   |  |    | |:          ||  |
Db    1434 rrsswectgsgraipftgsgkcpqkdhtstavgdgssgsqpregrg----digtncqkmt 1489

Qy     566 ENSVLSHSAK----VSPAEHDIQIMSDLHE--------QSLPKK--KKKQKLEVTREKQT 611
             |  ||||      ::|| | :|:  :           || |::  :: |:  :|  ::
Db    1490 -nktlshsesqvfaltpalhklqlgeetqsdepnleslqseprelpqrsqeanmtesrk- 1547

Qy     612 MIDDIPMDIVELLAKNQHERQLMTETDCSDINRIQSKTTADDDCVIVAAKDGSDY----- 666
                     |::: ::    :    : |:  :||  ||   :|  :: |
Db    1548 ------------aedemqksawdqpslpegnk--nktnldd---lvkgenrssvkhrla 1589

Qy     667 ----ASSVFDTN--SQQKSLAS---QSTQKELQGHLALTTQE------SPHPQNFQS--- 708
               ||  |     | :: :|:   ||  :    ||:| | |       | |:: :|
Db    1590 amskasrkfpakdvsprrhvatifpqsgsrsgfdhlslgtvecnplfpeptpksaesige 1649

Qy     709 --TQEQQTHLRMEE--------------MVTIAASSPLFSHHDDQY--IAEAPTEHWGR 749
               | : |::  |               :|:   |: :   | :::  ::|:|::|
Db    1650 srlsenrkhvkksenllpitvlpnrepsthvsnqksnsisqrhqnefknvsespskhens 1709

Qy     750 KDAKKLTWEQFKATTRNSPAATCGAQFRPGIQAVDLTSTHVMGSSSNYASRQPVIAPLDR 809
           ||          |   :    ||        :  ||     | |     |   ||:
Db    1710 kd----------vtaaqnlvresgap-----spitftslreaefsdnqrrlsppf-plep 1753

Qy     810 YAERAVNQVHARNFPSTIATMEASKLCDRRNAGQVVLYPKESMPATHLLRMMDPSTLASF 869
            |::           | :::  || |  :|:|  :     |  |  || |     :| |
Db    1754 -aqk----------srvssplasflqqqrsassl-----ewepephlyr---skslksi 1793

Qy     870 PNYGTSSRNQMESQLHNSQYAHNQYKGSTSTSYGSNLNGKIPLTFEDLSRHQLHDLHRPL 929
             :|   |       | :     :  |||     |   :: |  |    :      |
Db    1794 nvhgdllrks-----hppkvrerhfsestsi---dnalsrltlgnefsvnngysrrfrsf 1845

Qy     930 RPHPRVGVLGSLLQKEIANWSENCGTQSGYKLGVSTGITSHQMNRKEHFEALNSGMFSAK 989
             |     |:       :|:   ||::| :  :|              :  :: |:|  :
Db    1846 selpscd--gn------eswayrsgtktgprsaisi-----------yrpidygifgke 1885

Qy     990 WNALQLGSVSSSADFLSARNSIAQSWTRGKGKMVHPLDRFVRQDICITNKNP 1041
               | :|  |   |:       |: :  |        |:: |:    :||
Db    1886 qqlaflenvkrs---ltqgrlwkpsflknpg--------flkddl----rnp 1922


RESULT   3
ABG19656
ID   ABG19656 standard; Protein; 1982 AA.
XX
AC   ABG19656;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #19647.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS83843.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 50015; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
```

```
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   1982 AA;


  Query Match             4.0%;  Score 221;  DB 22;  Length 1982;
  Best Local Similarity   18.6%;  Pred. No. 1.5e-07;
  Matches  210;  Conservative  171;  Mismatches  423;  Indels  328;  Gaps   53;

Qy      83 HDQ---KKCDEHKASSSPFSVAKFRRWDCSKCLDKLKTSDNGTAPRTLPAKQNGTSDGCS 139
           ||:    ||  |     :     | |      |::||  ::: : |          |||  |
Db     946 hdevvdvkchshspfrnergkgkirhh--isciekIsktesisvp---------tsdhrs 994

Qy     140 ITFVRSTFVPASVGSQKVS---------PSTQSS---QGKNADRSTLPKSVQEG------ 181
           :       : |:  : |||          |   ||    |: :    :  |:: |
Db     995 l-------ieanqsnskvseldtiyctlprksssflihgrqsgskimaaslrngpppfqi 1047

Qy     182 ------------NDSKCNAPSGKNGAAEANTDSPMKDLQGPAQNYDVAANVSEDNTSVD 228
                       |    ::|   |  ::  :||    |: |     :    |  :|||
Db    1048 knnvedamgnymlnkfspsspesanecskvlsdsa---leapeatermtnvkssgstsvr 1104

Qy     229 VGALP-------EVPQITWHIE----------VNGADQPPSTPKLSEVVLKRNEDENGKT 271
           | ||         |    |           :| |    ||:  | ::   | ::
Db    1105 kgplpflinramscpsgephastgregrkkpltsgmdaseltpraweriispvesdss-- 1162

Qy     272 EETLVAEQCNLTK------------------------------DPNP----MSGKERDQ 296
                 |:|||                                || |    :|||||  :
Db    1163 -----vrdcsltkrqhqkenfqeytekegkmaasrrsvfalsnedplpfcsdlsgkergk 1217

Qy     297 VAEQCNLTKDPKPVSGQKCEQICNEPCEEVVLKRSSKSKRKTDKKLMKKQQHSKK----- 351
           :   |     ||| :     |  | |||    :  :: : |     ||:::
Db    1218 tlhkvk-ttstfsvsgde----dnvkclevvsiyytlprkpskkfcnllqqytqntnlli 1272

Qy     352 RTAQADVSDAKLCRRKPKK------------VRLLSEIINANQVEDSRSDEVHRENAADP 399
           : | :          | |:                |   :|::|   : :    | :   |
Db    1273 espqvetetfpnalekdkqnystreqsgtpscenlkmsvnsdqtlttenmtafrlsnrgp 1332

Qy     400 CEDD-------RSTIPVPMEVSMDIPVSNHTVGEDGLKSSKNKTKRKYSDVVDDGSSLMN 452
                     : : :| | |    : :  : :  |  |:|| :|
Db    1333 laptlqemasveaavslpeeeskareifsdnlaktplgdsenkker-------------- 1378

Qy     453 WLNGKKKRTGSVHHTVAHPAGNLSNKKVTPTASTQHDDENDTENGLDTNMHKTDVCQHVS 512
           |||  :: ::| ::     |:| :| :       :  |   : |   : : ::    :
Db    1379 ---gkklqsetlhtslmlqrknvseek-sencqqsinssnsgpsslpa-lsevnignsqt 1433

Qy     513 EISTQRCSSKGKT-----AGLSKGKTHSAASTKYG--GESTRNGQNIHVLSAEDQCQMET 565
             |:  |:  |:      :|    | |:: :   |  |   | |:         ||  |
Db    1434 rrsswectgsgraipftgsgkcpqkdhtstavgdgssgsqpregrg----digtncqkmt 1489

Qy     566 ENSVLSHSAK----VSPAEHDIQIMSDLHE--------QSLPKK--KKKQKLEVTREKQT 611
           |  ||||      ::|| | :|:  :            || |::  :: |:  :|  ::
Db    1490 -nktlshsesqvfaltpalhklqlgeetqsdepnleslqseprelpqrsqeanmtesrk- 1547

Qy     612 MIDDIPMDIVELLAKNQHERQLMTETDCSDINRIQSKTTADDDCVIVAAKDGSDY----- 666
                     |::: ::    :    : |:  :||  ||   :|  :: |
Db    1548 ------------aedemqksawdqpslpegnk--nktnldd---lvkgenrssvkhrla 1589

Qy     667 ----ASSVFDTN--SQQKSLAS---QSTQKELQGHLALTTQE------SPHPQNFQS--- 708
             ||  |      | :: :|:   ||  :    ||:| | |       | |:: :|
Db    1590 amskasrkfpakdvsprrhvatifpqsgsrsgfdhlslgtvecnplfpeptpksaesige 1649

Qy     709 --TQEQQTHLRMEE--------------MVTIAASSPLFSHHDDQY--IAEAPTEHWGR 749
             | : |::  |                |:   |: :   | :::  ::|:|::|
```

```
Db    1650 srlsenrkhvkksenllpitvlpnrepsthvsnqksnsisqrhqnefknvsespskhens 1709

Qy     750 KDAKKLTWEQFKATTRNSPAATCGAQFRPGIQAVDLTSTHVMGSSSNYASRQPVIAPLDR 809
           ||           |   :     ||       : ||      | |     |   ||:
Db    1710 kd----------vtaaqnlvresgap-----spitftslreaefsdnqrrlsppf-plep 1753

Qy     810 YAERAVNQVHARNFPSTIATMEASKLCDRRNAGQVVLYPKESMPATHLLRMMDPSTLASF 869
            |::           | :::  || |  :|:|  :     |  |  || |     :| |
Db    1754 -aqk----------srvssplasflqqqrsassl-----ewepephlyr---skslksi 1793

Qy     870 PNYGTSSRNQMESQLHNSQYAHNQYKGSTSTSYGSNLNGKIPLTFEDLSRHQLHDLHRPL 929
             :|   |       |  :     :  |||     |   :: |  |   :      |
Db    1794 nvhgdllrks-----hppkvrerhfsestsi---dnalsrltlgnefsvnngysrrfrsf 1845

Qy     930 RPHPRVGVLGSLLQKEIANWSENCGTQSGYKLGVSTGITSHQMNRKEHFEALNSGMFSAK 989
             |     |:       :|:   ||::|  :  :|                :  :: |:|  :
Db    1846 selpscd--gn------eswayrsgtktgprsaisi-----------yrpidygifgke 1885

Qy     990 WNALQLGSVSSSADFLSARNSIAQSWTRGKGKMVHPLDRFVRQDICITNKNP 1041
             | :|  |   |:       |: :  |       |:: |:    :||
Db    1886 qqlaflenvkrs---ltqgrlwkpsflknpg--------flkddl----rnp 1922


RESULT   4
ABB29725
ID   ABB29725 standard; Peptide; 1325 AA.
XX
AC   ABB29725;
XX
DT   01-FEB-2002  (first entry)
XX
DE   Peptide #2376 encoded by breast cell single exon nucleic acid probe.
XX
KW   Human; microarray; single exon probe; gene expression; breast;
KW   disease; cancer.
XX
OS   Homo sapiens.
XX
PN   WO200157271-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00662.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-496933/54.
XX
PT   New spatially-addressable set of single exon nucleic acid probes,
PT   useful for measuring gene expression in sample derived from human
PT   breast, comprises number of single exon nucleic acid probes -
XX
PS   Claim 27; SEQ ID NO 12693; 327pp + sequence listing; English.
XX
CC   The invention relates to a spatially-addressable set of single exon
CC   nucleic acid probes for measuring gene expression in a sample derived
CC   from human breast and BT 474 cells. The method involves contacting
CC   the probes with a collection of detectably labelled nucleic acids
CC   derived from mRNA of human breast, and then measuring the label
CC   bound to each probe of the microarray. The probes are useful for
CC   verifying the expression of regions of genomic DNA predicted to
CC   encode proteins. They are useful for gene discovery, and for
CC   determining predisposition and/or prognosing breast disease. Gene
```

```
CC   expression analysis is useful for assessing the toxicity of chemical
CC   agents on cells. The microarray of this invention presents a far greater
CC   diversity of probes for measuring gene expression, with far less bias
CC   than expressed sequence tag microarrays. The method is suitable for
CC   rapid production of functional information from genomic sequence. The
CC   present sequence is a peptide encoded by a single exon nucleic acid
CC   probe of the invention.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   1325 AA;


  Query Match             4.0%;  Score 220;  DB 22;  Length 1325;
  Best Local Similarity   19.4%;  Pred. No. 9.7e-08;
  Matches  153;  Conservative  125;  Mismatches  311;  Indels  198;  Gaps   38;

Qy     113 DKLKTSDNGTAPRTLPAKQNGTSDGCSITFVRSTFVPASVGSQKVSPSTQSSQGKNADRS 172
           |:   ::  | |      :||          |: |      |    |:    :   |: :
Db     536 drtplanekttpslaeptengq---------rtpfanekttsssaepteheertplanen 586

Qy     173 TLPKSVQEGNDSKCNAPSGKNGAAEANTDSPMKDLQGPAQNYDVAANVSEDNTSVDVGAL 232
           | | |  |  :::      :  | |  | ||     || :| :: ||   : |::
Db     587 ttp-spaeptenr------ertanenttpsp----agptenremtan---ekttlfpaep 632

Qy     233 PEVPQITWHIEVNGADQPPS-----TPKLSEVVLKRNEDENGKTEETLVAEQCNLTKDP- 286
             | : | : :   :   |:     ||  :|       :     | | :| : | |  |
Db     633 tenrertanekttsspaeptengqrtpfanekttsspaeptehgertplane-nttlspa 691

Qy     287 NPMSGKE-----------------RDQVAEQCNLTKDP---------KPVSGQKCEQICN 320
             |   :|                 |:: | : | |  |          |:: :|
Db     692 eptenrertanekttpfpaeptenrertane-nttpspaqptengdrtplanekttpsla 750

Qy     321 EPCEEVVLKRSSKSKRKTDKKLMKKQQHSKKRTAQADVSDAKLCRRKPKKVRLLSEIINA 380
           || |    ||:  :  ||     :  :|:::                       : : |
Db     751 epteng--krtpfanekttsssaeptehaer----------------------tplane 785

Qy     381 NQVEDSRSDEVHRENA---------ADPCEDDRS-----TIPVPMEVSMDIPVSNH--TV 424
           |          :||           |:| |:  |     | | | |     |  |   |
Db     786 nttsspaeptenrertanekttqfpaeptenrestanekttpfpae-----ptenrewta 840

Qy     425 GEDGLKSSKNKTKRKYSDVVDDGSSLMNWLNGKKKRTGSVHHTVAHPAGN-----LSNKK 479
             |:   |    |:       :: : | |     :| | |     | |  |      :|:|
Db     841 nenttlspaepteh------eemtplan-----ekttls----paeptengertpftnek 885

Qy     480 VTPTAS--TQHDDENDTENGLDTNMHKTDVCQHVSEISTQRC-------SSKGKTAGLSK 530
           ||:::  |:| :     | : |   : :  :|   |: ::        :  |:|  :::
Db     886 ttpssaeptehgertplanei-ttpsraeptehgerianekatpspakptehgett-vne 943

Qy     531 GKTHSAASTKYGG--------------ESTRNGQNIHVLSAEDQCQMETENSVLSHSAK 575
             | |:|     |                |||  :|:               ||:   : ||
Db     944 dttpssaeptengertplanentttsptestehgertanekttpspaeptehgertpsan 1003

Qy     576 -------VSPAEHDIQIMSDLHEQSLPKKKKK----QKLEVTREKQTMIDDIPMDIVELL 624
                 | ||: ::    :| : |   |     :|  :  || |:  : |    |
Db    1004 ektipspakptehe-emtpsanenttpspvkptehgekttlanekitlspegp---tehg 1059

Qy     625 AKNQHERQLMTETDCSDINRIQSKTTADDDCVIVAAKDGSDYASSVFDTN---------- 674
           ||     : :| :          :||: :| :  :| : :::       |
Db    1060 akttsanekitpslakpteh-gerttspndkitssaaestehrdratsanvitpapaepi 1118

Qy     675 --SQQKSLASQSTQKELQGHLALTTQESPHPQNFQSTQEQQTHLRMEEMVTIAASSPLFS 732
             ::: :|| :   :       :| : : ||:   || |: |  |  |  |: :   : :
Db    1119 khakrttlahekmtq-------vtekstehpekttsttektt--rtpekptlysektict 1169

Qy     733 HHDDQYIAEAPTEHWGRKDAKKLTWEQFKATTRN--SPAATCGA--QFRP-----GIQAV 783
             :  : | |||: |      || |  ||  ::  :|  |       :|     | :::
Db    1170 kgkntpvpekptenlgn---ttlttetikapvkstenpektaavtktikpsvkvtgdksl 1226

Qy     784 DLTSTHV 790
             ||:|:
Db    1227 tttsshl 1233
```

```
RESULT   5
ABB34897
ID   ABB34897 standard; Peptide; 1325 AA.
XX
AC   ABB34897;
XX
DT   04-FEB-2002  (first entry)
XX
DE   Peptide #2403 encoded by human foetal liver single exon probe.
XX
KW   Human; foetal liver; gene expression; single exon nucleic acid probe.
XX
OS   Homo sapiens.
XX
PN   WO200157277-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00669.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-483447/52.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human fetal liver -
XX
PS   Claim 27; SEQ ID NO 27532; 639pp + sequence listing; English.
XX
CC   The invention relates to a single exon nucleic acid probe for
CC   measuring human gene expression in a sample derived from human foetal
CC   liver. The single exon nucleic acid probes may be used for predicting,
CC   measuring and displaying gene expression in samples derived from human
CC   fetal liver. The present sequence is a peptide encoded by a single exon
CC   nucleic acid probe of the invention.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   1325 AA;


  Query Match             4.0%;  Score 220;  DB 22;  Length 1325;
  Best Local Similarity   19.4%;  Pred. No. 9.7e-08;
  Matches  153;  Conservative  125;  Mismatches  311;  Indels  198;  Gaps   38;

Qy     113 DKLKTSDNGTAPRTLPAKQNGTSDGCSITFVRSTFVPASVGSQKVSPSTQSSQGKNADRS 172
           |:   ::  | |      :||          |: |      |    |:    :   |: :
Db     536 drtplanekttpslaeptengq---------rtpfanekttsssaepteheertplanen 586

Qy     173 TLPKSVQEGNDSKCNAPSGKNGAAEANTDSPMKDLQGPAQNYDVAANVSEDNTSVDVGAL 232
           | | |  |  :::      :  | |  | ||     || :| :: ||   : |::
Db     587 ttp-spaeptenr------ertanenttpsp----agptenremtan---ekttlfpaep 632

Qy     233 PEVPQITWHIEVNGADQPPS-----TPKLSEVVLKRNEDENGKTEETLVAEQCNLTKDP- 286
             | : | : :   :   |:     ||  :|       :     | | :| : | |  |
Db     633 tenrertanekttsspaeptengqrtpfanekttsspaeptehgertplane-nttlspa 691

Qy     287 NPMSGKE-----------------RDQVAEQCNLTKDP---------KPVSGQKCEQICN 320
             |   :|                 |:: | : | |  |          |:: :|
Db     692 eptenrertanekttpfpaeptenrertane-nttpspaqptengdrtplanekttpsla 750
```

```
Qy    321 EPCEEVVLKRSSKSKRKTDKKLMKKQQHSKKRTAQADVSDAKLCRRKPKKVRLLSEIINA 380
Db    751 epteng--krtpfanekttsssaeptehaer----------------------tplane 785

Qy    381 NQVEDSRSDEVHRENA---------ADPCEDDRS-----TIPVPMEVSMDIPVSNH--TV 424
Db    786 nttsspaeptenrertanekttqfpaeptenrestanekttpfpae-----ptenrewta 840

Qy    425 GEDGLKSSKNKTKRKYSDVVDDGSSLMNWLNGKKKRTGSVHHTVAHPAGN-----LSNKK 479
Db    841 nenttlspaepteh------eemtplan-----ekttls----paeptengertpftnek 885

Qy    480 VTPTAS--TQHDDENDTENGLDTNMHKTDVCQHVSEISTQRC-------SSKGKTAGLSK 530
Db    886 ttpssaeptehgertplanei-ttpsraeptehgerianekatpspakptehgett-vne 943

Qy    531 GKTHSAASTKYGG--------------ESTRNGQNIHVLSAEDQCQMETENSVLSHSAK 575
Db    944 dttpssaeptengertplanentttsptestehgertanekttpspaeptehgertpsan 1003

Qy    576 -------VSPAEHDIQIMSDLHEQSLPKKKKK----QKLEVTREKQTMIDDIPMDIVELL 624
Db   1004 ektipspakptehe-emtpsanenttpspvkptehgekttlanekitlspegp---tehg 1059

Qy    625 AKNQHERQLMTETDCSDINRIQSKTTADDDCVIVAAKDGSDYASSVFDTN---------- 674
Db   1060 akttsanekitpslakpteh-gerttspndkitssaaestehrdratsanvitpapaepi 1118

Qy    675 --SQQKSLASQSTQKELQGHLALTTQESPHPQNFQSTQEQQTHLRMEEMVTIAASSPLFS 732
Db   1119 khakrttlahekmtq-------vtekstehpekttsttektt--rtpekptlysektict 1169

Qy    733 HHDDQYIAEAPTEHWGRKDAKKLTWEQFKATTRN--SPAATCGA--QFRP-----GIQAV 783
Db   1170 kgkntpvpekptenlgn---ttlttetikapvkstenpektaavtktikpsvkvtgdksl 1226

Qy    784 DLTSTHV 790
Db   1227 tttsshl 1233
```

RESULT 6
ABB20314
ID ABB20314 standard; Protein; 1325 AA.
XX
AC ABB20314;
XX
DT 23-JAN-2002 (first entry)
XX
DE Protein #2313 encoded by probe for measuring heart cell gene expression.
XX
KW Human; gene expression; heart; microarray; vascular system;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease.
XX
OS Homo sapiens.
XX
PN WO200157274-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00666.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MQLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488899/53.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT hearts -
XX
PS Claim 15; SEQ ID No 22084; 530pp; English.
XX
CC The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart (see
CC ABA21535-ABA41305). The present sequence is a protein encoded by one such
CC probe. The probes may be used for predicting, measuring and displaying
CC gene expression in samples derived from the human heart via microarrays.
CC By measuring gene expression, the probes are useful for predicting,
CC diagnosing, grading, staging, monitoring and prognosing diseases of the
CC human heart and vascular system e.g. cardiovascular disease,
CC hypertension, cardiac arrhythmias and congenital heart disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1325 AA;

Query Match 4.0%; Score 220; DB 22; Length 1325;
Best Local Similarity 19.4%; Pred. No. 9.7e-08;
Matches 153; Conservative 125; Mismatches 311; Indels 198; Gaps 38;

```
Qy    113 DKLKTSDNGTAPRTLPAKQNGTSDGCSITFVRSTFVPASVGSQKVSPSTQSSQGKNADRS 172
Db    536 drtplanekttpslaeptengq---------rtpfanekttsssaepteheertplanen 586

Qy    173 TLPKSVQEGNDSKCNAPSGKNGAAEANTDSPMKDLQGPAQNYDVAANVSEDNTSVDVGAL 232
Db    587 ttp-spaeptenr------ertanenttpsp----agptenremtan---ekttlfpaep 632

Qy    233 PEVPQITWHIEVNGADQPPS-----TPKLSEVVLKRNEDENGKTEETLVAEQCNLTKDP- 286
Db    633 tenrertanekttsspaeptengqrtpfanekttsspaeptehgertplane-nttlspa 691

Qy    287 NPMSGKE-----------------RDQVAEQCNLTKDP---------KPVSGQKCEQICN 320
Db    692 eptenrertanekttpfpaeptenrertane-nttpspaqptengdrtplanekttpsla 750

Qy    321 EPCEEVVLKRSSKSKRKTDKKLMKKQQHSKKRTAQADVSDAKLCRRKPKKVRLLSEIINA 380
Db    751 epteng--krtpfanekttsssaeptehaer----------------------tplane 785

Qy    381 NQVEDSRSDEVHRENA---------ADPCEDDRS-----TIPVPMEVSMDIPVSNH--TV 424
Db    786 nttsspaeptenrertanekttqfpaeptenrestanekttpfpae-----ptenrewta 840

Qy    425 GEDGLKSSKNKTKRKYSDVVDDGSSLMNWLNGKKKRTGSVHHTVAHPAGN-----LSNKK 479
Db    841 nenttlspaepteh------eemtplan-----ekttls----paeptengertpftnek 885

Qy    480 VTPTAS--TQHDDENDTENGLDTNMHKTDVCQHVSEISTQRC-------SSKGKTAGLSK 530
Db    886 ttpssaeptehgertplanei-ttpsraeptehgerianekatpspakptehgett-vne 943

Qy    531 GKTHSAASTKYGG--------------ESTRNGQNIHVLSAEDQCQMETENSVLSHSAK 575
Db    944 dttpssaeptengertplanentttsptestehgertanekttpspaeptehgertpsan 1003

Qy    576 -------VSPAEHDIQIMSDLHEQSLPKKKKK----QKLEVTREKQTMIDDIPMDIVELL 624
Db   1004 ektipspakptehe-emtpsanenttpspvkptehgekttlanekitlspegp---tehg 1059

Qy    625 AKNQHERQLMTETDCSDINRIQSKTTADDDCVIVAAKDGSDYASSVFDTN---------- 674
Db   1060 akttsanekitpslakpteh-gerttspndkitssaaestehrdratsanvitpapaepi 1118
```

```
Qy     675 --SQQKSLASQSTQKELQGHLALTTQESPHPQNFQSTQEQQTHLRMEEMVTIAASSPLFS 732
           ::: :|| :   :       :| : : ||:   || |: |  |  |  |: :   : :
Db    1119 khakrttlahekmtq-------vtekstehpekttsttektt--rtpekptlysektict 1169

Qy     733 HHDDQYIAEAPTEHWGRKDAKKLTWEQFKATTRN--SPAATCGA--QFRP-----GIQAV 783
           :  : | |||: |      || |  ||  ::  :|  |       :|     | :::
Db    1170 kgkntpvpekptenlgn---ttlttetikapvkstenpektaavtktikpsvkvtgdksl 1226

Qy     784 DLTSTHV 790
           ||:|:
Db    1227 tttsshl 1233


RESULT   7
AAM55707
ID   AAM55707 standard; Protein; 1325 AA.
XX
AC   AAM55707;
XX
DT   05-NOV-2001  (first entry)
XX
DE   Human brain expressed single exon probe encoded protein SEQ ID NO: 27812.
XX
KW   Human; brain expressed exon; gene expression analysis; probe;
KW   microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW   epilepsy; cancer.
XX
OS   Homo sapiens.
XX
PN   WO200157275-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00667.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-483446/52.
XX
PT   Single exon nucleic acid probes for analyzing gene expression in human
PT   brains -
XX
PS   Example 4; SEQ ID NO: 27812; 650pp + Sequence Listing; English.
XX
CC   The present invention provides a number of single exon nucleic acid
CC   probes which are derived from genomic sequences expressed in the human
CC   brain. They can be used to measure gene expression in brain cell samples,
CC   which may enable the diagnosis and improved treatment of nervous system
CC   diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC   epilepsy and cancers. The present sequence is a protein encoded by one of
CC   the probes of the invention.
XX
SQ   Sequence   1325 AA;


  Query Match             4.0%;  Score 220;  DB 22;  Length 1325;
  Best Local Similarity   19.4%;  Pred. No. 9.7e-08;
  Matches  153;  Conservative  125;  Mismatches  311;  Indels  198;  Gaps   38;

Qy     113 DKLKTSDNGTAPRTLPAKQNGTSDGCSITFVRSTFVPASVGSQKVSPSTQSSQGKNADRS 172
           |:   ::  | |      :||          |: |      |    |:    :   |: :
Db     536 drtplanekttpslaeptengq---------rtpfanekttsssaepteheertplanen 586

Qy     173 TLPKSVQEGNDSKCNAPSGKNGAAEANTDSPMKDLQGPAQNYDVAANVSEDNTSVDVGAL 232
           | | |  |  :::      :  | |  | ||     || :| :: ||   : |::
Db     587 ttp-spaeptenr------ertanenttpsp----agptenremtan---ekttlfpaep 632

Qy     233 PEVPQITWHIEVNGADQPPS-----TPKLSEVVLKRNEDENGKTEETLVAEQCNLTKDP- 286
           |  : | : :   :   |:     ||  :|       :     | | :| : | |  |
Db     633 tenrertanekttsspaeptengqrtpfanekttsspaeptehgertplane-nttlspa 691

Qy     287 NPMSGKE-----------------RDQVAEQCNLTKDP---------KPVSGQKCEQICN 320
           |   :|                 |:: | : | |  |          |:: :|
Db     692 eptenrertanekttpfpaeptenrertane-nttpspaqptengdrtplanekttpsla 750

Qy     321 EPCEEVVLKRSSKSKRKTDKKLMKKQQHSKKRTAQADVSDAKLCRRKPKKVRLLSEIINA 380
           || |    ||:  :  ||     :  :|:::                    : : |
Db     751 epteng--krtpfanekttsssaeptehaer----------------------tplane 785

Qy     381 NQVEDSRSDEVHRENA---------ADPCEDDRS-----TIPVPMEVSMDIPVSNH--TV 424
           |          :||           |:| |:  |     | | | |     |  |   |
Db     786 nttsspaeptenrertanekttqfpaeptenrestanekttpfpae-----ptenrewta 840

Qy     425 GEDGLKSSKNKTKRKYSDVVDDGSSLMNWLNGKKKRTGSVHHTVAHPAGN-----LSNKK 479
           |:   |    |:        :: : | |     :| | |     | |  |       :|:|
Db     841 nenttlspaepteh------eemtplan-----ekttls----paeptengertpftnek 885

Qy     480 VTPTAS--TQHDDENDTENGLDTNMHKTDVCQHVSEISTQRC-------SSKGKTAGLSK 530
           ||:::  |:| :     | : |   : :  :|   |: ::        :  |:|  :::
Db     886 ttpssaeptehgertplanei-ttpsraeptehgerianekatpspakptehgett-vne 943

Qy     531 GKTHSAASTKYGG---------------ESTRNGQNIHVLSAEDQCQMETENSVLSHSAK 575
             | |:|     |                |||  :|:              ||:   : ||
Db     944 dttpssaeptengertplanentttsptestehgertanekttpspaeptehgertpsan 1003

Qy     576 -------VSPAEHDIQIMSDLHEQSLPKKKKK----QKLEVTREKQTMIDDIPMDIVELL 624
                 | ||: ::    :| : |   |     :|  :  || |:  : |    |
Db    1004 ektipspakptehe-emtpsanenttpspvkptehgekttlanekitlspegp---tehg 1059

Qy     625 AKNQHERQLMTETDCSDINRIQSKTTADDDCVIVAAKDGSDYASSVFDTN---------- 674
           ||     : :| :           :||: :| :  :| : :::       |
Db    1060 akttsanekitpslakpteh-gerttspndkitssaaestehrdratsanvitpapaepi 1118

Qy     675 --SQQKSLASQSTQKELQGHLALTTQESPHPQNFQSTQEQQTHLRMEEMVTIAASSPLFS 732
             ::: :|| :   :       :| : : ||:   || |: |  |  |  |: :   : :
Db    1119 khakrttlahekmtq-------vtekstehpekttsttektt--rtpekptlysektict 1169

Qy     733 HHDDQYIAEAPTEHWGRKDAKKLTWEQFKATTRN--SPAATCGA--QFRP-----GIQAV 783
           :  : | |||: |      || |  ||  ::  :|  |       :|     | :::
Db    1170 kgkntpvpekptenlgn---ttlttetikapvkstenpektaavtktikpsvkvtgdksl 1226

Qy     784 DLTSTHV 790
           ||:|:
Db    1227 tttsshl 1233


RESULT   8
AAM68085
ID   AAM68085 standard; Protein; 1325 AA.
XX
AC   AAM68085;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human bone marrow expressed probe encoded protein SEQ ID NO: 28391.
XX
KW   Human; bone marrow expressed exon; gene expression analysis; probe;
KW   microarray; cancer; leukaemia; lymphoma; myeloma.
XX
OS   Homo sapiens.
XX
PN   WO200157276-A2.
XX
PD   09-AUG-2001.
XX
```

```
PF   30-JAN-2001; 2001WO-US00668.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488900/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human bone marrow -
XX
PS   Example 4; SEQ ID NO: 28391; 658pp + Sequence Listing; English.
XX
CC   The present invention provides a number of single exon nucleic acid
CC   probes which are derived from genomic sequences expressed in the human
CC   bone marrow. They can be used to measure gene expression in bone marrow
CC   samples, which may enable the improved diagnosis and treatment of cancers
CC   such as lymphoma, leukaemia and myeloma. The present sequence is a
CC   protein encoded by one of the probes of the invention.
XX
SQ   Sequence   1325 AA;

  Query Match          4.0%;  Score 220;  DB 22;  Length 1325;
  Best Local Similarity   19.4%;  Pred. No. 9.7e-08;
  Matches  153;  Conservative  125;  Mismatches  311;  Indels  198;  Gaps   38;

Qy     113 DKLKTSDNGTAPRTLPAKQNGTSDGCSITFVRSTFVPASVGSQKVSPSTQSSQGKNADRS 172
           |:   ::  | |      :||          |: |      |    |:    :   |: :
Db     536 drtplanekttpslaeptengq---------rtpfanekttsssaepteheertplanen 586

Qy     173 TLPKSVQEGNDSKCNAPSGKNGAAEANTDSPMKDLQGPAQNYDVAANVSEDNTSVDVGAL 232
           | | |  |  :::      :  | |  | ||     || :| :: ||   : |::
Db     587 ttp-spaeptenr------ertanenttpsp----agptenremtan---ekttlfpaep 632

Qy     233 PEVPQITWHIEVNGADQPPS-----TPKLSEVVLKRNEDENGKTEETLVAEQCNLTKDP- 286
           |  : | : :   :   |:     ||  :|       :     | | :| : | |  |
Db     633 tenrertanekttsspaeptengqrtpfanekttsspaeptehgertplane-nttlspa 691

Qy     287 NPMSGKE----------------RDQVAEQCNLTKDP---------KPVSGQKCEQICN 320
           |   :|                |:: | : | |  |         |:: :|
Db     692 eptenrertanekttpfpaeptenrertane-nttpspaqptengdrtplanekttpsla 750

Qy     321 EPCEEVVLKRSSKSKRKTDKKLMKKQQHSKKRTAQADVSDAKLCRRKPKKVRLLSEIINA 380
           || |    ||:  :  ||     :  :|:::                      : : |
Db     751 epteng--krtpfanekttsssaeptehaer----------------------tplane 785

Qy     381 NQVEDSRSDEVHRENA---------ADPCEDDRS-----TIPVPMEVSMDIPVSNH--TV 424
           |          :||           |:| |:  |     | | | |     |  |   |
Db     786 nttsspaeptenrertanekttqfpaeptenrestanekttpfpae-----ptenrewta 840

Qy     425 GEDGLKSSKNKTKRKYSDVVDDGSSLMNWLNGKKKRTGSVHHTVAHPAGN-----LSNKK 479
           |:   |    |:       :: : | |     :| | |     | |  |      :|:|
Db     841 nenttlspaepteh------eemtplan-----ekttls----paeptengertpftnek 885

Qy     480 VTPTAS--TQHDDENDTENGLDTNMHKTDVCQHVSEISTQRC-------SSKGKTAGLSK 530
           ||:::  |:| :     | : |   : :  :|   |: ::        :  |:|  :::
Db     886 ttpssaeptehgertplanei-ttpsraeptehgerianekatpspakptehgett-vne 943

Qy     531 GKTHSAASTKYGG--------------ESTRNGQNIHVLSAEDQCQMETENSVLSHSAK 575
           | |:|     |               ||| :|:              ||:   : ||
Db     944 dttpssaeptengertplanentttsptestehgertanekttpspaeptehgertpsan 1003

Qy     576 -------VSPAEHDIQIMSDLHEQSLPKKKKK----QKLEVTREKQTMIDDIPMDIVELL 624
                 | ||: ::    :| : |   |     :|  :  || |:  : |    |
Db     1004 ektipspakptehe-emtpsanenttpspvkptehgekttlanekitlspegp---tehg 1059

Qy     625 AKNQHERQLMTETDCSDINRIQSKTTADDDCVIVAAKDGSDYASSVFDTN---------- 674
           ||     : :| :          :||: :| :  :| : :::       |
Db     1060 akttsanekitpslakpteh-gerttspndkitssaaestehrdratsanvitpapaepi 1118

Qy     675 --SQQKSLASQSTQKELQGHLALTTQESPHPQNFQSTQEQQTHLRMEEMVTIAASSPLFS 732
             ::: :|| :   :       :| : : ||:   || |: |  |  |  |: :   : :
Db     1119 khakrttlahekmtq------vtekstehpekttsttektt--rtpekptlysektict 1169

Qy     733 HHDDQYIAEAPTEHWGRKDAKKLTWEQFKATTRN--SPAATCGA--QFRP-----GIQAV 783
             :  : | |||: |      || |  ||  ::  :|  |       :|     | :::
Db     1170 kgkntpvpekptenlgn---ttlttetikapvkstenpektaavtktikpsvkvtgdksl 1226

Qy     784 DLTSTHV 790
             ||:|:
Db     1227 tttsshl 1233


RESULT   9
AAM03645
ID   AAM03645 standard; Protein; 1325 AA.
XX
AC   AAM03645;
XX
DT   09-OCT-2001  (first entry)
XX
DE   Peptide #2327 encoded by probe for measuring breast gene expression.
XX
KW   Probe; human; breast disease; breast cancer; development disorder;
KW   inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
OS   Homo sapiens.
XX
PN   WO200157270-A2.
XX
PD   09-AUG-2001.
XX
PF   29-JAN-2001; 2001WO-US00661.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-476286/51.
XX
PT   Novel single exon nucleic acid probe used to measuring gene expression
PT   in a human breast -
XX
PS   Claim 27; SEQ ID No 12385; 322pp; English.
XX
CC   The present invention relates to novel single exon nucleic acid probes
CC   (see AAI00010-AAI10067). The present sequence is a peptide encoded by one
CC   such probe. The probes are useful for measuring human gene expression in
CC   a human breast sample, where the probe hybridises at high stringency to a
CC   nucleic acid expressed in the human breast. The probes are useful for
CC   predicting, diagnosing, grading, staging, monitoring and prognosing
CC   diseases of the human breast, particularly those diseases with polygenic
CC   aetiology. The diseases include: breast cancer, disorders of development,
CC   inflammatory diseases of the breast, fibrocystic changes, proliferative
CC   breast disease and non-carcinoma tumours.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
```

```
SQ   Sequence   1325 AA;

  Query Match             4.0%;  Score 220;  DB 22;  Length 1325;
  Best Local Similarity   19.4%;  Pred. No. 9.7e-08;
  Matches  153;  Conservative  125;  Mismatches  311;  Indels  198;  Gaps   38;

Qy     113 DKLKTSDNGTAPRTLPAKQNGTSDGCSITFVRSTFVPASVGSQKVSPSTQSSQGKNADRS 172
Db     536 drtplanekttpslaeptengq---------rtpfanekttsssaepteheertplanen 586

Qy     173 TLPKSVQEGNDSKCNAPSGKNGAAEANTDSPMKDLQGPAQNYDVAANVSEDNTSVDVGAL 232
Db     587 ttp-spaeptenr------ertanenttpsp----agptenremtan---ekttlfpaep 632

Qy     233 PEVPQITWHIEVNGADQPPS-----TPKLSEVVLKRNEDENGKTEETLVAEQCNLTKDP- 286
Db     633 tenrertanekttsspaeptengqrtpfanekttsspaeptehgertplane-nttlspa 691

Qy     287 NPMSGKE----------------RDQVAEQCNLTKDP---------KPVSGQKCEQICN 320
Db     692 eptenrertanekttpfpaeptenrertane-nttpspaqptengdrtplanekttpsla 750

Qy     321 EPCEEVVLKRSSKSKRKTDKKLMKKQQHSKKRTAQADVSDAKLCRRKPKKVRLLSEIINA 380
Db     751 epteng--krtpfanekttsssaeptehaer----------------------tplane 785

Qy     381 NQVEDSRSDEVHRENA---------ADPCEDDRS-----TIPVPMEVSMDIPVSNH--TV 424
Db     786 nttsspaeptenrertanekttqfpaeptenrestanekttpfpae-----ptenrewta 840

Qy     425 GEDGLKSSKNKTKRKYSDVVDDGSSLMNWLNGKKKRTGSVHHTVAHPAGN-----LSNKK 479
Db     841 nenttlspaepteh------eemtplan-----ekttls----paeptengertpftnek 885

Qy     480 VTPTAS--TQHDDENDTENGLDTNMHKTDVCQHVSEISTQRC-------SSKGKTAGLSK 530
Db     886 ttpssaeptehgertplanei-ttpsraeptehgerianekatpspakptehgett-vne 943

Qy     531 GKTHSAASTKYGG---------------ESTRNGQNIHVLSAEDQCQMETENSVLSHSAK 575
Db     944 dttpssaeptengertplanentttsptestehgertanekttpspaeptehgertpsan 1003

Qy     576 -------VSPAEHDIQIMSDLHEQSLPKKKKK----QKLEVTREKQTMIDDIPMDIVELL 624
Db    1004 ektipspakptehe-emtpsanenttpspvkptehgekttlanekitlspegp---tehg 1059

Qy     625 AKNQHERQLMTETDCSDINRIQSKTTADDDCVIVAAKDGSDYASSVFDTN---------- 674
Db    1060 akttsanekitpslakpteh-gerttspndkitssaaestehrdratsanvitpapaepi 1118

Qy     675 --SQQKSLASQSTQKELQGHLALTTQESPHPQNFQSTQEQQTHLRMEEMVTIAASSPLFS 732
Db    1119 khakrttlahekmtq-------vtekstehpekttsttektt--rtpekptlysektict 1169

Qy     733 HHDDQYIAEAPTEHWGRKDAKKLTWEQFKATTRN--SPAATCGA--QFRP-----GIQAV 783
Db    1170 kgkntpvpekptenlgn---ttlttetikapvkstenpektaavtktikpsvkvtgdksl 1226

Qy     784 DLTSTHV 790
Db    1227 tttsshl 1233


RESULT  10
AAG70791
ID   AAG70791 standard; Protein; 1165 AA.
XX
AC   AAG70791;
XX
DT   27-JUL-2001  (first entry)
XX
DE   S cerevisiae apoptosis associated protein YMR230W.
XX
KW   Yeast; fungus; apoptosis; infection; proliferative disease;
KW   vaccine; autoimmune disease; ischaemia; neurodegeneration.
XX
OS   Saccharomyces cerevisiae.
XX
PN   WO200102550-A2.
XX
PD   11-JAN-2001.
XX
PF   03-JUL-2000; 2000WO-BE00077.
XX
PR   01-JUL-1999;   99EP-0870141.
XX
PA   (JANC ) JANSSEN PHARM NV.
XX
PI   Contreras RH,  De Backer MD,  Luyten WHML,  Malcorps IKL;
PI   Nelissen BJM,  Reekmans RJ;
XX
DR   WPI; 2001-367042/38.
DR   N-PSDB; AAH29827.
XX
PT   Yeast and fungal nucleic acids encoding proteins involved in a pathway
PT   leading to programmed cell death, useful for treating proliferative
PT   disorders, yeast and fungal infections, or for preventing apoptosis in
PT   certain diseases -
XX
PS   Claim 1; Fig 1; 218pp; English.
XX
CC   The present invention provides the protein and coding sequences of a
CC   number of apoptosis associated proteins from the yeast Saccharomyces
CC   cerevisiae and the fungus Candida albicans. These can be used to identify
CC   treatments for fungal and yeast infections, for proliferative diseases
CC   and for apoptosis related diseases such as autoimmune diseases, ischaemia
CC   and neurodegeneration. The present sequence is one of the S. cerevisiae
CC   proteins of the invention.
XX
SQ   Sequence   1165 AA;

  Query Match             3.9%;  Score 217;  DB 22;  Length 1165;
  Best Local Similarity   18.7%;  Pred. No. 1.4e-07;
  Matches  204;  Conservative  156;  Mismatches  416;  Indels  314;  Gaps   46;

Qy     155 QKVSPSTQSSQGKNADRSTLPKSVQEGNDSKCNAPSGKNGAAEANTDSPM--KDLQGPAQ 212
Db       6 rkltvetetveapvannlll------snnsnvvapnpsipsastst-splhreivddsva 58

Qy     213 NYDVAANVSE------DNTSVDVGALPEVPQITWHIEVNGADQPPSTPKLSEVVLKRNED 266
Db      59 tanttsnvvqhnlptidnnlmdsdatsh-nqdhwhsdinragtsmst---sdiptdlhle 114

Qy     267 ENGKTEETLVAEQCNLTKDPNPMSGKERDQVAEQCNLTKDPKPVSGQKCEQICNEPCEEV 326
Db     115 higsvsst-nnnsnnalinhnplssh---------------------------------- 139

Qy     327 VLKRSSKSKRKTDKKLMKKQQHSKKRTAQADVSDAKLCRRKPKKVRLLSEIINANQVEDS 386
Db     140 -lsnpssslrnkkssllvasn-------pafasdvelskkkpavisnnmptsnialyqta 191

Qy     387 RSDEVHRENAADPCEDDR------------------STIPVPMEVSMDIPVSNHTVGED 427
Db     192 rsanihgpsstsaskafrkasafsnntapstsnnigsntppapl-----lplps------ 240

Qy     428 GLKSSKNKTK---RKYSDVVDDGSSLM--NWLNGKKKRTGSVHHTVAHPAGNLSNKKV-T 481
Db     241 --lsqqnkpkiierptmhvtnsreillgenllddtkaknapanst-thdngpvandglri 297

Qy     482 PTASTQHDDENDTENGLDTNMH-----KTDVCQHVSEISTQRCSS----------KGKTA 526
Db     298 pnhsnaddnennnkmkknknsingknernddtskicttstktapstaplgstdntqalta 357

Qy     527 GLSKGKTHSAASTKYGGESTRNGQNIHVLSAEDQCQMETENSVLSHSAKVSPAEHDIQIM 586
```

```
Db    358 svsssnadnhnnnkkktssnnngnnsnsasnktnadiknsnadlsastsnnnaind---- 413

Qy    587 SDLHEQSLPKKKKKQ----------KLEVTREKQTMIDDIPMDIVELL-------AKNQ 628

Db    414 -dshesnsekptkadffaarlatavgeneisdseetfvyesaanstknlifpdsssqqqq 472

Qy    629 HERQLMTETDCSDINRIQSKTTA---DDDCVIVAAKDGSDYAS---------SVFDTNSQ 676

Db    473 qqqqppkqqqqqqnhgitskisapllnnnkkllsrlknsrhistgailnntiatistnpn 532

Qy    677 QKSLASQSTQKELQGH-----LALTTQESPHPQNFQSTQEQQTHLRMEEMVTIAASSPLF 731

Db    533 lnsnvmqnnnnlmsghnhldelssikqepph-----qlqqqqppmdvqsvdsytsdnp-- 585

Qy    732 SHHDDQYIAEAPTEHWGRKDAKKLTWEQFKATTRNSPAATC------GAQFRPGIQAVDL 785

Db    586 ---dsnviakspdkrsslvslskvsphllsstssngntiscpnvatnsqelepn---ndi 639

Qy    786 TSTHVMGSS----------SNYASRQ--------------PVIAPLDRYA--------ER 813

Db    640 stkkslsnstlrhssanrnsnygdnkrplrttvskifdsnpngaplrrysgvpdhvnled 699

Qy    814 AVNQVHARNFPSTIATMEASKLCDRRNAGQVVLYPKESMPATHLLRMMDPSTLASFPNYG 873

Db    700 yieqph--nyptmqnsvkkdefynsrn---------nkfp--hglnfygdnnvieeennq 746

Qy    874 TSS---------------------------------RNQMESQLHNSQYAH-------- 891

Db    747 dssnvnrpqhtnlqhefipednesdendihsmfyynhkndletkplisdygededvddyd 806

Qy    892 ------NQYKGSTSTSYGSNLNGKIPL----TFEDL-------SRHQLHDLHRPLRPHPR 934

Db    807 rpnatfnsyygsasnthelplhgrmpsrsnndyydfmvgnntgnnnqlne-ytplrmkrg 865

Qy    935 VGVLGSLLQKEIANWSENCG-----TQSGYKLGVSTGITSHQM-NRKEHFEALNSGMFSA 988

Db    866 qrhl-srtnnsimngsihmngnddvthsninnndivgysphnfysrkspfvkvk------ 918

Qy    989 KWNALQLGSVSSS-------ADFLSARNSIAQSWTRGKGKMVHPLDRFVRQ------DIC 1035

Db    919 --nflylafvissllmtgfilgfllatnkelqdvd------vvvmdnvisssdelifdit 970

Qy   1036 ITNKNPADFT 1045

Db    971 vsafnpgffs 980
```

RESULT 11
AAW20056
ID AAW20056 standard; Protein; 1528 AA.
XX
AC AAW20056;
XX
DT 10-SEP-1997 (first entry)
XX
DE C. elegans UNC-53 protein variant 8A.
XX
KW UNC-53; neuronal regeneration; revascularisation; wound healing;
KW neurodegenerative disease; Alzheimer's disease; Huntingdon's;
KW peripheral neuropathies; metastasis inhibition; cancer.
XX
OS Caenorhabditis elegans.
XX
PN WO9638555-A2.
XX
PD 05-DEC-1996.
XX
PF 31-MAY-1996; 96WO-EP02311.
XX
PR 31-MAY-1995; 95GB-0010944.
XX
PA (BOGA/) BOGAERT T.
PA (STRI/) STRINGHAM E.
PA (VAND/) VANDEKERCKHOVE J.
XX
PI Bogaert T, Stringham E, Vandekerckhove J;
XX
DR WPI; 1997-034369/03.
DR N-PSDB; AAT71314.
XX
PT Caenorhabditis elegans UNC-53 protein 8A and 7A variants - useful to
PT promote neuronal regeneration, revascularisation or wound healing.
XX
PS Claim 20; Page 106-111; 278pp; English.
XX
CC UNC-53 protein variants 8A and 7A of Caenorhabditis elegans are new.
CC The UNC-53 proteins and nucleic acids are useful as medicaments to
CC promote neuronal regeneration, revascularisation or wound healing, or
CC for treatment of chronic neurodegenerative diseases (e.g. Alzheimer's or
CC Huntingdon's disease) or acute traumatic injuries. Transgenic cells and
CC organisms transfected with UNC-53 cDNA can be used to determine whether
CC a substance is an inhibitor or enhancer of the regulation of cell shape
CC or motility or the direction of cell migration by screening for a
CC phenotypic change in the cell. Inhibitors can be used to alleviate the
CC spread of disease inducing cells or metastasis. Probes derived from the
CC cDNA sequences can be used to identify homologues of the C. elegans
CC unc-53 gene. The UNC-53 protein can be used to identify proteins which
CC are active in the signal transduction pathway that can be used as
CC mentioned above.
XX
SQ Sequence 1528 AA;

Query Match 3.8%; Score 208; DB 18; Length 1528;
Best Local Similarity 19.5%; Pred. No. 1.1e-06;
Matches 231; Conservative 174; Mismatches 427; Indels 350; Gaps 57;

```
Qy     14 VGTNCM------LARGGTGAVAPVLELTATPRQDAAAEAGVDEPAQHQCEHFSIRGYVAL 67

Db     83 lgldcskltktdidsgnlgavlqllfllstykq-------------------------- 115

Qy     68 LQKKDPKFCSLSRIFHDQKKCDEHKASSSPFSVAKFRRWDCSKCLDKLKTSDNGTAPRTL 127

Db    116 ---------klrqlkkdqkkleqlptsimppavsklpsprvatsatasatnpnsnfpqms 166

Qy    128 PAK-QNGTSDGCSITFVRSTFVPASVGSQKVSPSTQSSQGKNADRSTLPKSVQEGND--- 183

Db    167 tsrlqtpqsriskidsskigikpktsglkppsssttssnntnsfr---pssrssgnnnvg 223

Qy    184 -----SKCNAPSGKNGAAEANTDSPMKDLQGPAQNYDVAANVSEDNTSVDVGALPEVPQI 238

Db    224 stistsaksslessstyssisnlnrptsqlqkpsrpqtqlvrva---tttkigs------- 273

Qy    239 TWHIEVNGADQPPSTPKLSEVV---LKRNEDENGKTEETLVAEQCNLTKDP-----NPMS 290

Db    274 ----sklaapkavstpklasvktigakqepdnsgggggmlklklfssknpssssnspqp 329

Qy    291 GKERDQVAEQCNLTKDPKPV-SGQKCEQICNEPCEEVVLKRSSKSK--------RKTDKK 341

Db    330 trkaaavpqqqtlskiaapvksglk-------pptsklgsatsmsklctpkvsyrktdap 382

Qy    342 LMKKQQHSKKRTAQADVSDAKLCRRKPKK----VRLLSEIINANQVEDSRSDEVHRENAA 397

Db    383 iisqq------------dskrcsksseeesgyagfnstsptssstegsls--mhstssk 427

Qy    398 DPCEDDRSTIPVPMEVSMDI-----------PVSNHTVGED-------GLKSSKNKTKRK 439

Db    428 sstsdekspssddltlnasivtairqpiaatpvspniinkpveekptlavkgvkstakkd 487

Qy    440 YSDVVDDGSSLMNWLNGKKKRTGSVHHTVAHPAGNLSNKKVT--PTASTQHDDENDTENG 497

Db    488 pppavpprdt--------qptigvvspimah-------kkltndpvisekpepeklqsms 532

Qy    498 LDTNMHKTDV--CQHVSEISTQRCSSKGKTAG----LSKGKTHSAASTKYGGESTRNGQN 551
```

```
Db    533 idt----tdvpplpplksvvplkmtsirqpptydvllkqgkitspvks-fgyeqs----- 582

Qy    552 IHVLSAEDQCQMETENSVLSH-SAKVSP-----AEHDIQIMSDLHEQSLPKKKKKQKLEV 605
                  :|:|:::| ||:|:|       | :       |: : | |  :
Db    583 -----------sasedsivahasaqvtpptktsgnhsl-------errmgknktsessgy 624

Qy    606 T-----------REKQTMIDDIPMDIVELLAKNQH----ERQLMTETDCSDINRIQSKTT 650
          |           |||    ||:        |:| :    |      :  || | :   :|
Db    625 tsdagvamcakmreklkeyddmtr-----raqngypdnfedssslssgisdnnelddist 679

Qy    651 AD---DDCVIVAAKDGSDYASSV-FDTNSQQK----SLASQSTQKELQGHLALTTQESPH 702
            |    |   ||:|  |||:  |   |:|  |    | :| |     :         ||: :
Db    680 ddlsgvdmatvaskh-sdyshfvrhptsssskprvpsrsstsvdsrsrae-----qenvy 733

Qy    703 PQNFQSTQEQQTHLRMEEMVTIAASSPLFSHH---DDQYIAEAPTEHWGRKDAKKLTWEQ 759
            |    |:          ||::  |  |      | : :|  |     | | |
Db    734 kllsqcrtsqrg----------aaatstfgqhslrspgyssysp--hls-vsadkdtmsm 780

Qy    760 FKATTR--NSPAATCGAQFRPGIQAVDLTSTHVMGSSSNYASRQPVIAPLDRYAERAVNQ 817
             |:|  :|   :   ||     ::|    |:   :|       :::|
Db    781 hsqtsrrpssqkpsysgqf----hsld-rkchlqeftstehrmaallsp----------- 824

Qy    818 VHARNFPSTIATMEASKLCDRRNAG--QVVLYPKESMPATHLLRMMDPSTLASFPNYGTS 875
            |  |::::  ::|     |: |     :| :       | |: |  : |       |
Db    825 ---rrvpnsmskydssgsysarsrggsstgiyge----tfqlhrlsdekspah------s 871

Qy    876 SRNQMESQLHNSQYAHNQYKGSTSTSYGSNLNGKIPLTFEDLSRHQLHDLHRPL------ 929
          ::::| |||           :::|:||| || |         | : |: | |
Db    872 aksemgsql----------slasttaygs-lneky--------ehairdmardlecyknt 912

Qy    930 -----RPHPRVGVLGSLLQKEIANWSENCGTQS------------GYKLGVSTGITSHQ 971
                :    | |  | ::::   :::    :             :   :|  : |:
Db    913 vdsltkkqenygalfdlfeqklrkltqhidrsnlkpeeairfrqdiahlrdisnhlasns 972

Qy    972 MNRKE---------------HFEALNSGMFSAKWNALQLGSVSSSADFLSARNSIAQSW 1015
            :  |               |  :::|   |:|   : | |   :           :||
Db    973 ahanegagellrqpslesvashrssmsssskssksqekislssfgkn----------kksw 1022

Qy   1016 TRGKGKMVHPLDRFVRQDICITNKN--PADFTTISNDNEYMD 1055
            |       | :| ::     |||   |   :||     :|
Db   1023 irss------lskftkk----knknydeahmpsisgsqgtld 1054


RESULT  12
AAW20057
ID   AAW20057 standard; Protein; 1583 AA.
XX
AC   AAW20057;
XX
DT   10-SEP-1997  (first entry)
XX
DE   C. elegans UNC-53 protein variant 7A.
XX
KW   UNC-53; neuronal regeneration; revascularisation; wound healing;
KW   neurodegenerative disease; Alzheimer's disease; Huntingdon's;
KW   peripheral neuropathies; metastasis inhibition; cancer.
XX
OS   Caenorhabditis elegans.
XX
PN   WO9638555-A2.
XX
PD   05-DEC-1996.
XX
PF   31-MAY-1996;   96WO-EP02311.
XX
PR   31-MAY-1995;   95GB-0010944.
XX
PA   (BOGA/) BOGAERT T.
PA   (STRI/) STRINGHAM E.
PA   (VAND/) VANDEKERCKHOVE J.
XX
PI   Bogaert T,  Stringham E,  Vandekerckhove J;
XX
DR   WPI; 1997-034369/03.
DR   N-PSDB; AAT71315.
XX
PT   Caenorhabditis elegans UNC-53 protein 8A and 7A variants - useful to
PT   promote neuronal regeneration, revascularisation or wound healing.
XX
PS   Claim 22; Page 111-116; 278pp; English.
XX
CC   UNC-53 protein variants 8A and 7A of Caenorhabditis elegans are new.
CC   The UNC-53 proteins and nucleic acids are useful as medicaments to
CC   promote neuronal regeneration, revascularisation or wound healing, or
CC   for treatment of chronic neurodegenerative diseases (e.g. Alzheimer's or
CC   Huntingdon's disease) or acute traumatic injuries. Transgenic cells and
CC   organisms transfected with UNC-53 cDNA can be used to determine whether
CC   a substance is an inhibitor or enhancer of the regulation of cell shape
CC   or motility or the direction of cell migration by screening for a
CC   phenotypic change in the cell. Inhibitors can be used to alleviate the
CC   spread of disease inducing cells or metastasis. Probes derived from the
CC   cDNA sequences can be used to identify homologues of the C. elegans
CC   unc-53 gene. The UNC-53 protein can be used to identify proteins which
CC   are active in the signal transduction pathway that can be used as
CC   mentioned above.
XX
SQ   Sequence   1583 AA;


  Query Match             3.8%;  Score 208;  DB 18;  Length 1583;
  Best Local Similarity   19.5%;  Pred. No. 1.1e-06;
  Matches  231;  Conservative  174;  Mismatches  427;  Indels  350;  Gaps   57;

Qy      14 VGTNCM------LARGGTGAVAPVLELTATPRQDAAAEAGVDEPAQHQCEHFSIRGYVAL 67
              :| :|       :  |  |||  :| | :| :|
Db      83 lgldcskltktdidsgnlgavlqllfllstykq-------------------------- 115

Qy      68 LQKKDPKFCSLSRIFHDQKKCDEHKASSSPFSVAKFRRWDCSKCLDKLKTSDNGTAPRTL 127
                     | ::  |||| ::   |  | :|:|      :       |: |   |:
Db     116 ---------klrqlkkdqkkleqlptsimppavsklpsprvatsatasatnpnsnfpqms 166

Qy     128 PAK-QNGTSDGCSITFVRSTFVPASVGSQKVSPSTQSSQGKNADRSTLPKSVQEGND--- 183
           :: |   |    |   :    | : | :  | || ||   |: |   | |   ||:
Db     167 tsrlqtpqsriskidsskigikpktsglkppsssttssnntnsfr---pssrssgnnnvg 223

Qy     184 -----SKCNAPSGKNGAAEANTDSPMKDLQGPAQNYDVAANVSEDNTSVDVGALPEVPQI 238
                  |  :  |    :: :| : |   || |::       |:   |:  :|:
Db     224 stistsakslessstyssisnlnrptsqlqkpsrpqtqlvrva---tttkigs------- 273

Qy     239 TWHIEVNGADQPPSTPKLSEVV---LKRNEDENGKTEETLVAEQCNLTKDP-----NPMS 290
                 | :  |||||: |      |:  | :|     ::  :   :|:|     :|
Db     274 ----sklaapkavstpklasvktigakqepdnsgggggmlklklfssknpssssnspqp 329

Qy     291 GKERDQVAEQCNLTKDPKPV-SGQKCEQICNEPCEEVVLKRSSKSK--------RKTDKK 341
           ::   | :|  |:|   || || |       |    :   :| ||          ||||
Db     330 trkaaavpqqqtlskiaapvksglk-------pptsklgsatsmsklctpkvsyrktdap 382

Qy     342 LMKKQQHSKKRTAQADVSDAKLCRRKPKK----VRLLSEIINANQVEDSRSDEVHRENAA 397
           :: :|             |:| | :  ::         |    ::  | | |  :|  ::
Db     383 iisqq------------dskrcsksseeesgyagfnstsptssstegsls--mhstssk 427

Qy     398 DPCEDDRSTIPVPMEVSMDI-----------PVSNHTVGED-------GLKSSKNKTKRK 439
               |::|     : ::  |           ||| : : :          :|  |:  |:
Db     428 sstsdekspssddltlnasivtairqpiaatpvspniinkpveekptlavkgvkstakkd 487

Qy     440 YSDVVDDGSSLMNWLNGKKKRTGSVHHTVAHPAGNLSNKKVT--PTASTQHDDENDTENG 497
              |    :         :   | |   :||        ||:|  |  | : : |
Db     488 pppavpprdt--------qptigvvspimah-------kkltndpvisekpepeklqsms 532

Qy     498 LDTNMHKTDV--CQHVSEISTQRCSSKGKTAG----LSKGKTHSAASTKYGGESTRNGQN 551
           :||    |||      :   :   : :|  :        | :||  |   : :| | :
Db     533 idt----tdvpplpplksvvplkmtsirqpptydvllkqgkitspvks-fgyeqs----- 582

Qy     552 IHVLSAEDQCQMETENSVLSH-SAKVSP-----AEHDIQIMSDLHEQSLPKKKKKQKLEV 605
                  :|:|:::| ||:|:|       | :       |: : | |  :
Db     583 -----------sasedsivahasaqvtpptktsgnhsl-------errmgknktsessgy 624
```

```
Qy    606 T----------REKQTMIDDIPMDIVELLAKNQH----ERQLMTETDCSDINRIQSKTT 650
Db    625 tsdagvamcakmreklkeyddmtr-----raqngypdnfedssslssgisdnneldist 679

Qy    651 AD---DDCVIVAAKDGSDYASSV-FDTNSQQK----SLASQSTQKELQGHLALTTQESPH 702
Db    680 ddlsgvdmatvaskh-sdyshfvrhptsssskprvpsrsstsvdsrsrae-----qenvy 733

Qy    703 PQNFQSTQEQQTHLRMEEMVTIAASSPLFSHH---DDQYIAEAPTEHWGRKDAKKLTWEQ 759
Db    734 kllsqcrtsqrg----------aaatstfgqhslrspgyssysp--hls-vsadkdtmsm 780

Qy    760 FKATTR--NSPAATCGAQFRPGIQAVDLTSTHVMGSSSNYASRQPVIAPLDRYAERAVNQ 817
Db    781 hsqtsrrpssqkpsysgqf----hsld-rkchlqeftstehrmaallsp---------- 824

Qy    818 VHARNFPSTIATMEASKLCDRRNAG--QVVLYPKESMPATHLLRMMDPSTLASFPNYGTS 875
Db    825 ---rrvpnsmskydssgsysarsrggsstgiyge----tfqlhrlsdekspah------s 871

Qy    876 SRNQMESQLHNSQYAHNQYKGSTSTSYGSNLNGKIPLTFEDLSRHQLHDLHRPL------ 929
Db    872 aksemgsql----------slasttaygs-lneky--------ehairdmardlecyknt 912

Qy    930 -----RPHPRVGVLGSLLQKEIANWSENCGTQS------------GYKLGVSTGITSHQ 971
Db    913 vdsltkkqenygalfdlfeqklrkltqhidrsnlkpeeairfrqdiahlrdisnhlasns 972

Qy    972 MNRKE---------------HFEALNSGMFSAKWNALQLGSVSSSADFLSARNSIAQSW 1015
Db    973 ahanegagellrqpslesvashrssmsssskssksqekislssfgkn----------kksw 1022

Qy   1016 TRGKGKMVHPLDRFVRQDICITNKN--PADFTTISNDNEYMD 1055
Db   1023 irss------lskftkk----knknydeahmpsisgsqgtld 1054
```

RESULT 13
ABG03703
ID ABG03703 standard; Protein; 2070 AA.
XX
AC ABG03703;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #3694.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS67890.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 34062; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2070 AA;

Query Match 3.7%; Score 206.5; DB 22; Length 2070;
Best Local Similarity 19.1%; Pred. No. 2.3e-06;
Matches 231; Conservative 188; Mismatches 509; Indels 279; Gaps 53;

```
Qy      1 MEIVA----VDQEGARVVGTNCMLARGGTGAVAPVLELTATPRQDAAAEAGVDEPAQHQC 56
Db    920 veiiaepcrveqdssdtmqttgllkgqglttllsdlakkknpqkssslsdq-mdhp----- 973

Qy     57 EHFSIRGYVALLQKKDPKFCSLSRIFH----------DQKKCDEH--------------- 91
Db    974 -dfs---------senpkivdssvnlhpkqelllmnnddrdppqhhsclpdqevingsli 1023

Qy     92 --KASSSPFSVA--KFRRWDCSKCLDKLKTSDNGTAPRTLPAKQNGTSDGC-SITFVRST 146
Db   1024 ngrqadspmstssgsrsfsvasmlpettred---------vtsnattntcdsctfvetn 1074

Qy    147 FVPASVGSQKVSPSTQSSQGKNADRSTLPKSVQEGNDSKCNAPSGKNGAAEANTDSPMKD 206
Db   1075 xysssсsksyfxp------gepxegeswppgxykgssskpseasllegdppfksqipkes 1128

Qy    207 LQGPAQNYDVAANVSEDNTSVDVGALP----EVPQITWHIEVNGA------DQPPSTPKL 256
Db   1129 gtgqae-----atpnefnsqgsieatmerplekpscslgiktsnaslqdstsqppsitsl 1183

Qy    257 SEVVLKRNED----------ENGKTEETLVAEQCNLTKDPNPMSGKERDQVAEQCNLTKD 306
Db   1184 svnnlihqssishplascaglsptseqttvpatvnltvssssygsqppgtisdvlni--- 1240

Qy    307 PKPVSGQKCEQICNEPCEEVVLKRSSKSKRKTDKKLMKKQQHSKKRTAQAD--VSDAKL- 363
Db   1241 pknsxipmtstipnsqiqepllkpshesrk----------dsakravqddlllssakrq 1289

Qy    364 --CRRKPKKVRLLSEIINANQVEDSRSDEVHRENAADPCEDDRSTIPVPMEVSMDIPVSN 421
Db   1290 khcqpaplrlesmslm---srtpdtisdqtqmmvsqippnssnsv----------vpvsn 1336

Qy    422 HTVGEDGL-----KSSKNKTKRKYSDVVDDGSSLMNWLNGKKKRTGSVHHTVAHPAGNLS 476
Db   1337 pahg-dgltrlfppsnnfvtpalrqtevqcgsqps---vaeqqqtqasqhlqasaaacss 1392

Qy    477 NKKVTPTAST----QHDDENDTENGLDTNMHKTDVCQH----------------VSEIST 516
Db   1393 srcispsxstistxsssssssssssnsnnnkqgsxergityikcsimylmqralstlsp 1452

Qy    517 QRCSSKGKTAGLSKGKTHSAAS-----------TKYGGESTRN-----------GQNI 552
```

```
Db   1453 imstnrgsatgssdaekeescsghpdlsavpynqstmgtdqsrsktgqphphhqqmqqqm 1512

Qy    553 HVLSAEDQCQMETENSVLSHSAKVSPAEHDIQIMSDLHEQSLPKKKKKQKLEVTREKQTM 612

Db   1513 qqhfgssqtekscenpstsrnhhnhpqnh---lnqdimhqqqdvgsrqqgsgvsseh--v 1567

Qy    613 IDDIPMDIVELLAKNQHERQLMTE----TDCSDI---------NRIQS--------KTTA 651

Db   1568 sghnpmq--rlltsrgleqqmvsqpsivtrssdmtctphrpernrvssysaealigktss 1625

Qy    652 DDDCVIVAAKDGSDYAS-----SVFDTNSQQKSLASQSTQKELQGHLALTTQESPHPQNF 706

Db   1626 nseqrmgisiqgsrvsdqlemrsyldv-prnkslaihnmqgrvd-htvasdirlsdcqtf 1683

Qy    707 Q---STQEQQTHLRMEEMVTIAASSPLFSHHDDQYIAEAPTEHWGRKD---AKKLTWEQF 760

Db   1684 kpsgasqqpqsnfevqssrnneignpvsslrsmqsqafrisqntgpppidrqkrlsyppv 1743

Qy    761 KA--TTRNSPAA----TCGAQFRPGIQAVDLTSTHVMGSSSNYASRQPVIAPLDRYAERA 814

Db   1744 qsiptgngipsrdsentchqsfmqsllaphl-sdqvigsqrslsehq-------rntqcg 1795

Qy    815 VNQVHARNFPSTIATMEASKLCDRRNAGQVVLYPKESMPATHLLRMMDPSTLASFPNYGT 874

Db   1796 pssaieyncppthenvhirresesqn--------rescdms-lgaintrnstlnipfsss 1846

Qy    875 SSRNQMESQLHN---SQYAHNQYKGSTSTSYGSNLNGKIPLTFEDLSRHQLHDLHRPLRP 931

Db   1847 sssgdiqgrntspnvsvqksnpmriteshatkghmnp--pvt------tnmhgvarpalp 1898

Qy    932 HPRV----GVLGSLLQKEIANWSENCGTQSGYKLGVSTGITSHQMNRKEHFEALNSGMFS 987

Db   1899 hpsvshgngdqgpavrqa----nssvpqrsrhplqdssgskirqpernrsgnqrqstvfd 1954

Qy    988 AKWNALQLGSVSSSADFLSARNSIAQSWTRGKGKMVHPLDRFVRQDICITNKNPADFTTI 1047

Db   1955 pslphlpl---stggsmilgrq---qpatekrgsiv----rfm---------?--pdspqv 1993

Qy   1048 SNDNEYM 1054

Db   1994 pndnsdm 2000


RESULT 14
ABB68397
ID   ABB68397 standard; Protein; 2768 AA.
XX
AC   ABB68397;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 31983.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL12500.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 31983; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   2768 AA;


  Query Match             3.7%;  Score 204;  DB 22;  Length 2768;
  Best Local Similarity   19.7%;  Pred. No. 5.6e-06;
  Matches  181;  Conservative  118;  Mismatches  345;  Indels  274;  Gaps   40;

Qy      29 APVLELTATPRQDAAAEAGVD--------------------EPAQHQCEHFSIRGYVALL 68

Db    1550 apvavttaapskddgeqkpveveekpiedgqkpiedetstptsseneiepesdrattiap 1609

Qy      69 QKKDPKFCSLSRIFHDQKKCDEHKASSSPFSVAKFRRWDCSKCLDKLKTSDNGTAPRTL- 127

Db    1610 skeepsepstgaptkde---paepstdapes------------desketpesevpttva 1653

Qy     128 PAKQNGTSDGCSITFVRSTFVPASVGSQKVSPSTQSSQGKNADRSTLPKSVQEGNDSKCN 187

Db    1654 pagek--------------iptssitpdeeptatsapvakpdedv---eketsteiptd 1695

Qy     188 APSGKNGAAEANTDS-----PMKDLQGPAQ---NYDVAANVSEDNTSVDV--------GA 231

Db    1696 apasseedensstdqipsevpekkpetpaqtpeegdivgataapttsdevppvqrlpeev 1755

Qy     232 LPEVPQITWHIEVNGADQPPSTPKL------------SEVVLKRNEDENGKTEETLVAE 278

Db    1756 laeipqpstetgikqqdettaapsidrkepyvteideeattvapisekdekpteeekpve 1815

Qy     279 QCNLTKDPNPMSGKERDQVAEQCNLTKDPKPVSGQKCEQICNEPCEEVVLKRSSKSKRKT 338

Db    1816 qkptgeeps--eeeekekpieqdvsteg--pvsteaseagstesseev--------kpst 1863

Qy     339 DKKLMKKQQHSK-KRTAQADVS-----DAKLCRR---KPKKVRLLSEIINANQVEDSRSD 389

Db    1864 egevaekpedkqpsstaqapvetipeistelpaqdgdkptseapvdsdedtsapsdekip 1923

Qy     390 EVHRENAADP---------CEDDRSTIPVPMEVSM--DIPVSNHTVGEDGLKSSKNKTKR 438

Db    1924 svsgeevegpevttaspqaaeedelktpaesepsstdkvpeteyqkpedetka--detpe 1981

Qy     439 KYSDVVDDGSSLMNWLNGKKKRTGSVHH----TVAHPAGNLSNKKVTPTASTQHD----- 489

Db    1982 svtqvsdvatstsapvag-----gdiekdeqattaspe---eeeeikptiapaaeipqps 2033

Qy     490 -----DENDTENGLDTNMHKTDVCQHVSEISTQRCSSKGKTAG-LSKGKTHSAASTKYGG 543

Db    2034 ekepvdeqevesgtkatpaesd-gqpideiap-------atsgpideast--aaptk--- 2080

Qy     544 ESTRNGQNIHVLSAEDQCQMETENSVLSHSAKVSPAEHDIQI---------MSDLHEQSLP 595

Db    2081 ----------------------eesttvasa-aspavhddeikdvtttqpvadekevaap 2117

Qy     596 KKKKKQKLEVTREKQTMIDDIPMDIVELLAKNQHERQLMTETDCSDINRIQSKTTADDDC 655
```

```
Db   2118 qdetktsidvstdsptaqdd-------------ekqdkteapvaptt--vssptadsaa 2161

Qy    656 ------------VIVAAKDGSDYASSVFDTNSQQKSLASQSTQKELQGHLALTTQE---- 699
                      | :  |   |  |     ::      || ||: | |  : :: |:
Db   2162 dsstptvevpspveidtkpmddimsqtiaphtad-gaaststededqapvtvspqdaekt 2220

Qy    700 --SPHPQNFQSTQEQQTHLRMEEMVTIAASSPLFSHHDDQYIAEAPTEHWGRKDAKKLTW 757
            || ||:   |   :     :|:   | ::||    ||
Db   2221 pvspapqdsdktpsseapqdadeip--atatpl----dd---------------------- 2253

Qy    758 EQFKATTRNSPAATCGAQFRPGIQAVDLTSTHVMGSSSNYASRQPVIAPLDRYAERAVNQ 817
                |   ||   |   |: |          :  :    |   ||||
Db   2254 --------nkipatvapqtddgvpat--------aapldedkiqttaaplde-------- 2289

Qy    818 VHARNFPSTIATMEASKL 835
              ||| | ::  |:
Db   2290 ---ekipstaaplddeki 2304


RESULT 15
ABB60291
ID   ABB60291 standard; Protein; 2951 AA.
XX
AC   ABB60291;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 7665.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL04394.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 7665; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   2951 AA;

  Query Match             3.7%;  Score 204;  DB 22;  Length 2951;
  Best Local Similarity   20.1%;  Pred. No. 6.2e-06;

  Matches  213;  Conservative  161;  Mismatches  463;  Indels  222;  Gaps   48;

Qy     78 LSRIFHDQKKCDEHKASSSPFSVAKFRRWDCS-------KCL-------DK-LKTSDNGT 122
          :|| |  :|  | |  |:|||    | :| :        ||       || |:|::|
Db   1838 vsrfthsaekvkepkv-sapfs----refdvnakipenddcllsinqkvdkflrtaenvi 1892

Qy    123 APRTLPAK------------QNGTSDGCSITFVRST--FVPASVGSQKVSPSTQSSQGK- 167
           | :|||::             :    | |:::  :    |:     ::|::|:      :
Db   1893 rptslpsrpeierpgleeideellrddctlsvsqkvhkfidt---aeklaptmpqksprl 1949

Qy    168 --NADRSTLPKSVQEGNDSKCNAPSGKNGAAEANTDSPMKDLQGPAQ---NYDVAANVSE 222
            | :|   :|  |    : : |       | : |: ::|   :|     ::   |::
Db   1950 vanierhisrqsepereldeesep-------eldrdtdvedddqtsqleteeeitqtvtk 2002

Qy    223 DNTSVDVGALPEVPQITWHIEVNGADQPPSTPKLSEVVLKRNEDENGKTEETLVAE-QCN 281
            |      | |  | |         |      ||     |::     | |   | : |
Db   2003 ket------lkefkqqtketretrrdskaepeklq----kkspqtkvkeesarvpkyqak 2052

Qy    282 LTKDPNPMSGKERDQVAEQCNLTKDPKPVSGQKCEQICNEPCEEVVLKRSSK-SKRKTDK 340
            :::  :    |:: |   :    : |      |   :: :||  | | ||  |  ::::
Db   2053 vsqkvsqwepkkqpqrepkvtqketplepkkqplskvkdep--ekvnkrepkvpqkesqt 2110

Qy    341 KLMKKQQHSKKRTAQADVSDAKLCRRKPKKVRLLSEIINANQVEDSRSDEVHRENAADPC 400
            || :: :   |:| |          ::|:|  |        | ||        |   :|
Db   2111 klkeepervtkktpq----------keprkepl-------rqsedepefspeeefddepl 2153

Qy    401 EDDRS-TIPVPMEVSMDIPVSNHTVGE--DGLKSSKNKTKRKYSDVVDDGSSLMNWLNGK 457
             :: |  : |:   ||      |:      |||   :  | :      |  | : :
Db   2154 pmtkthttaiemkrqkdilnrpsvfgqrtperkssttpsptklngtrgrpspstnlitee 2213

Qy    458 KKRTGSVHHTVAHPAGNLSNKKVTPTASTQHDDENDTENGLDTNMHK-----------T 505
            |:   :    |: |      :| ||:|::           :   | :
Db   2214 krsyrnqvtnvskp----gtrkttpsanspaqspppkttsiskrmeqisqqswvvqdvdv 2269

Qy    506 DV-------CQHVSEISTQRCSSKGKTAGLSKGKTHS---AASTKYGGESTRNGQNIHVL 555
            ||         |:||    :  |   :  ||:  : |   : :| |    | |    :|
Db   2270 dvevvgpappshisekpqgkspsptssrslsrspsrspkksittty----ttnttgrnva 2325

Qy    556 SAEDQCQMETENSVLSHSAKVSPAEHDIQIMSDLHEQSLPKKKKKQKLEVTR----EKQT 611
            |  :   |  | |       |:  |       | :|  :  ::  ||: :    | :
Db   2326 srrnvfepvhethvdseptgrrpsymd-------htksslehirrdsleinkshysrkss 2378

Qy    612 MIDDIPMDIVELLAKNQHERQLMTETDCSDINRIQSKTTADD--DCVIVAAKDGSDY--- 666
            | || |::      :| :           |: |  |   ||:      :    : ||
Db   2379 meddspve-----prnpnssvkf------dvprksssrgadeprktslkgkdedsdlele 2427

Qy    667 ASSVFDTNSQQK---SLASQSTQKELQGHLALTTQESPHPQNFQSTQEQQTHLRMEEMVT 723
             :||    :|   ::||   :: ::  : |        :|  :     |   :    |
Db   2428 ieeifdlqrlekllletvasyemrrriraqmrlir------knminagttttittittstt 2481

Qy    724 IAASSPLFSHHDDQYIAEAPTEHWGRKDAKKLTWEQFKATTRNSPAATCGAQFRPGIQAV 783
             ||||     ||  | |                 :  |:      |   | :  :: |
Db   2482 pgkssplpkirrdqspagaa----------------evktkevrtttsrrqqqqrveqv 2524

Qy    784 DLTSTHVMGSSSNYASRQPVIAPLDRYAERAVNQVHARNFP---------STIATMEASK 834
            | |:    | :|  :  :| : | :| |  |  :   |  |          ||  | : :
Db   2525 dsttpiapgktsphg--kppvkprersaspaqkr---risppgkqspgdrsttttktvtt 2579

Qy    835 LCDRRNAGQVVLYPKESMPA--------THLLR--MMDPSTLASFPNYGTSSRNQMESQL 884
             | |          | ||        : :|:     | |  | |  |::|     |
Db   2580 tsttrga--------pskpaqgpiwadrskvlkghatvpqtngstprkgstssttsss-- 2629

Qy    885 HNSQYAHNQYKGSTSTSYGSNLNGKIPLTFEDLSRHQLHDLHRPLRPHPRVGV----LGS 940
             :          ||:||  |  | :    ||            : |    |:    : :
Db   2630 --gkitrtmtsssttttsssttntnrnkqreedsitssy-----gvgptdenglplfgira 2682

Qy    941 LLQKEIANWSENCGTQ---SGYKLG---VSTGITSHQMNRKEHFEALNSGMFSAKWNALQ 994
            | :|       | | |:   :|| :      |   :  :  |||   : |:  ':|    ||
Db   2683 lkkkatppaeepcetkqevtgyvieeqfysdnksppрherkeliyssnadelaaikqqlq 2742

Qy    995 LGSVSSSADFLSARNSIAQSWTRGKGKMVHPLD-RFVRQ 1032
             ||   | ||  : : : : : :   | | |:||:
Db   2743 -deddsspplldar--vvrefkkvesqqslpedaryvrr 2778
```

Search completed: September 16, 2002, 22:23:16
Job time: 8005 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: September 16, 2002, 20:19:56 ; Search time 32.9 Seconds
(without alignments)
784.737 Million cell updates/sec

Title: US-09-828-068-2
Perfect score: 5526
Sequence: 1 MEIVAVDQEGARVVGTNCML.....NKNPADFTTISNDNEYMDYR 1057

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	202.5	3.7	1848	4	US-08-296-791-6	Sequence 6, Appli
2	202.5	3.7	1848	5	PCT-US95-10661A-6	Sequence 6, Appli
3	182	3.3	900	1	US-08-425-061-20	Sequence 20, Appl
4	182	3.3	900	2	US-08-825-886-20	Sequence 20, Appl
5	180	3.3	1863	2	US-08-603-753D-2	Sequence 2, Appli
6	180	3.3	1863	4	US-09-099-753-2	Sequence 2, Appli
7	180	3.3	1863	4	US-08-986-106-2	Sequence 2, Appli
8	180	3.3	1863	4	US-09-007-678B-49	Sequence 49, Appl
9	179	3.2	1202	1	US-08-425-061-22	Sequence 22, Appl
10	179	3.2	1202	2	US-08-825-886-22	Sequence 22, Appl
11	179	3.2	1363	1	US-08-425-061-23	Sequence 23, Appl
12	179	3.2	1363	2	US-08-825-886-23	Sequence 23, Appl
13	179	3.2	1852	1	US-08-425-061-24	Sequence 24, Appl
14	179	3.2	1852	2	US-08-825-886-24	Sequence 24, Appl
15	179	3.2	1863	1	US-08-480-784-2	Sequence 2, Appli
16	179	3.2	1863	1	US-08-483-553-2	Sequence 2, Appli
17	179	3.2	1863	1	US-08-487-002-2	Sequence 2, Appli
18	179	3.2	1863	1	US-08-483-554B-2	Sequence 2, Appli
19	179	3.2	1863	1	US-08-488-011B-2	Sequence 2, Appli
20	179	3.2	1863	4	US-08-850-727-2	Sequence 2, Appli
21	179	3.2	1863	5	PCT-US95-10202-2	Sequence 2, Appli
22	179	3.2	1863	5	PCT-US95-10203-2	Sequence 2, Appli
23	179	3.2	1863	5	PCT-US95-10220-2	Sequence 2, Appli
24	178	3.2	914	1	US-08-425-061-21	Sequence 21, Appl
25	178	3.2	914	2	US-08-825-886-21	Sequence 21, Appl
26	178	3.2	1863	1	US-08-598-591-2	Sequence 2, Appli
27	178	3.2	1863	1	US-08-798-691-2	Sequence 2, Appli
28	178	3.2	1863	1	US-08-798-691-4	Sequence 4, Appli
29	178	3.2	1863	1	US-08-798-691-6	Sequence 6, Appli
30	178	3.2	1863	3	US-08-825-487A-2	Sequence 2, Appli
31	178	3.2	1863	3	US-08-825-487A-4	Sequence 4, Appli
32	178	3.2	1863	3	US-08-825-487A-6	Sequence 6, Appli
33	178	3.2	1863	4	US-09-074-476-2	Sequence 2, Appli
34	178	3.2	1863	4	US-09-074-476-4	Sequence 4, Appli
35	178	3.2	1863	4	US-09-074-476-6	Sequence 6, Appli
36	172	3.1	1863	1	US-08-425-061-16	Sequence 16, Appl
37	172	3.1	1863	2	US-08-825-886-16	Sequence 16, Appl
38	168.5	3.0	1805	1	US-07-853-913-2	Sequence 2, Appli
39	166.5	3.0	1780	1	US-08-769-309A-5	Sequence 5, Appli
40	166.5	3.0	1780	3	US-08-994-570-5	Sequence 5, Appli
41	160.5	2.9	1589	3	US-08-755-587-189	Sequence 189, App
42	159	2.9	688	3	US-09-141-047-8	Sequence 8, Appli
43	158	2.9	976	4	US-08-894-997-50	Sequence 50, Appl
44	158	2.9	2285	4	US-09-308-375-2	Sequence 2, Appli
45	157.5	2.9	1093	5	PCT-US93-03077-1	Sequence 1, Appli

ALIGNMENTS

```
RESULT   1
US-08-296-791-6
; Sequence 6, Application US/08296791
; Patent No. 6245337
;  GENERAL INFORMATION:
;    APPLICANT:  St. Geme III, Joseph W.
;    APPLICANT:  Falkow, Stanley
;    TITLE OF INVENTION:  Haemophilus Adherence and Penetration
;    TITLE OF INVENTION:  Protein
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/296,791
;      FILING DATE:  25-AUG-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Trecartin, Richard F.
;      REGISTRATION NUMBER:  31,801
;      REFERENCE/DOCKET NUMBER:  A-59941/RFT/RMS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1848 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
US-08-296-791-6


  Query Match             3.7%;  Score 202.5;  DB 4;  Length 1848;
  Best Local Similarity   18.5%;  Pred. No. 8.7e-07;
  Matches  177;  Conservative  141;  Mismatches  403;  Indels  235;  Gaps   34;

Qy    130 KQNGTSDGCSITFVRSTFVPASVGSQKVSPSTQSSQGKNADRST-LPKSVQEGNDSKCNA 188
          |:| | |  :||       |  :  |  :|| ||:  : |   | :|           |
Db    999 KRNQTVDTTNIT------TPNDI--QADAPSAQSNNEEIARVETPVPPP----------A 1040
```

```
Qy     189 PSGKNGAAEANTDSPMKDLQGPAQNYDVAANVSEDNTSVDVGALPEVPQITWHIEVNGAD 248
           |: ::  |    ::   :   ||      || :|     |     | |        |
Db    1041 PATESAIASEQPETRPAETAQPAMEETNTANSTETAPKSDTATQTENP--------NSES 1092

Qy     249 QPPSTPKLSEVVLKRNEDENGKTEETLVAEQCNLTKDPNPMSGK--ERDQVAEQCNLTKD 306
             |  |   :| | :    ||    ||:   :   | :| | :|:  : ||   : |   :
Db    1093 VPSET---TEKVAENPPQEN----ETVAKNEQEAT-EPTPQNGEVAKEDQPTVEANTQTN 1144

Qy     307 PKPVSGQKCE-----QICNEPCEEVVLKRSSKSKRKT----DKKLMKKQQHSKKRTAQAD 357
             | | |      :  :|| | | :  :   |  :    || :::|:: :|  | :
Db    1145 EATQSEGKTEETQTAETKSEPTESVTVSENQPEKTVSQSTEDKVVVEKEEKAKVETEETQ 1204

Qy     358 VSDAKLCRRKPKKVRLLSEIINANQVEDSRSDEVHRENAADPCEDDRSTIPVPMEVSMDI 417
           :     :  ||:              :   :||  :  |:  :  : | |  :  : :
Db    1205 KAPQVTSKEPPKQA------------EPAPEEVPTDTNAEEAQALQQTQPTTV-AAAET 1250

Qy     418 PVSNHTVGEDGLKSSKNKTKRKYSDVVDDGSSLM-----NWLNGKKKRTGSVHHTVAHPA 472
             |    |:  : |:       : || : ::            :|::|  |    :  :
Db    1251 TSPNSKPAEETQQPSEKTNAEPVTPVVSENTATQPTETEETAKVEKEKTQEVPQVASQES 1310

Qy     473 GNLSNKKVTPTASTQHDDENDTENGLDTNMHKTDVCQHVSEISTQRCSSKGKTAGLSKGK 532
                 | | |:   |   || | |         |   :  :    : |:||  ||
Db    1311 PKQEQPAAKPQAQTKPQAEPARENVLTTKNVGEPQPQAQPQTQSTAVPTTGETAANSKPA 1370

Qy     533 THSAASTKYGGESTRNGQNIHVLSAEDQCQMETENSVLSHSAKVSPAEHDIQIMSDLHEQ 592
             |  |   |  |  :|:  :: ::  | :|       || ||  :   :  |:: ||
Db    1371 AKPQAQAKPQTEPAR--ENVSTVNTKEP-QSQT-------SATVSTEQPAKETSSNV-EQ 1419

Qy     593 SLPKKKKKQKLEVTREKQTMIDDIPMDIVELLAKNQHERQLMTETDCSDINRIQSKTTAD 652
           |:         |  :     | |   :| : :|  : :  |  | |  |  :|
Db    1420 PAPENSINTGSATTMTETAEKSDKPQ--METVTENDRQPEANTVADNSVANNSESS---- 1473

Qy     653 DDCVIVAAKDGSDYASSVFDTNSQQKSLASQSTQKELQGHLALTTQE--------SPHPQ 704
                          :: |:::   ||  :   :     :|||        :| |:
Db    1474 ------------------ESKSRRRRSVSQPKETSAEETTVASTQETTVDNSVSTPKPR 1514

Qy     705 NFQSTQEQQTHLRMEEMVTIAASSPLFSHHDDQYIAEAPTEHWGRKDAKKLTWEQFKATT 764
           : :: :  ||:          :  |:          | |||:
Db    1515 SRRTRRSVQTN----------SYEPV----------ELPTEN------------------ 1536

Qy     765 RNSPAATCGAQFRPGIQAVDLTSTHVMGSSSNYASRQPVIAPLDRYAERAVNQVHARNFP 824
                 |:    :|           | :| |: ||           |:  : ::|
Db    1537 ---------AENAENVQ----------SGNNVANSQP----------ALRNLTSKNTN 1565

Qy     825 STIATMEASKLCDRRNAGQVVLYPKESMPATHLLRMMDPSTLASFPNYGT----SSRNQM 880
           : |:   |       | |: |       : |: ::          | |       |   |
Db    1566 AVISNAMAKAQFVALNVGKAV--------SQHISQL-------EMNNEGQYNVWISNTSM 1610

Qy     881 ESQLHNSQYAHNQYKGSTSTSYGSNLNGKIPLTFEDLSRHQLHDLHRPLRPHPRVGVLGS 940
                | |:  ||:  :| | : |     ::    :  ||  :  :|          |
Db    1611 -----NKNYSSEQYRRFSSKSTQTQLGWDQTIS----NNVQLGGVFTYVRNSNNFDKASS 1661

Qy     941 LLQKEIANWSENCGTQSGYKLGVSTGITSHQMNRKEH---------------FEALNSG 984
                 |:       : :||:  |    | | : :                  :| | |
Db    1662 KNTLAQVNFYSKYYADNHWYLGIDLGYGKFQSNLQTNNNAKFARHTAQIGLTAGKAFNLG 1721

Qy     985 MFSAKWNALQLGSVSSSADFLSARNSIAQSWTRGKGKMVHPLD---RFVRQDICIT 1037
           |: |       |  |:|||  |:: |           |:|:     | : |:  |
Db    1722 NFAVKPTVGVRYSYLSNADFALAQDRI----------KVNPISVKTAFAQVDLSYT 1767
```

```
RESULT   2
PCT-US95-10661A-6
; Sequence 6, Application PC/TUS9510661A
;  GENERAL INFORMATION:
;    APPLICANT:  Washington University, et al.
;    TITLE OF INVENTION:  Haemophilus Adherence and Penetration Protein
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/10661A
;      FILING DATE:  16-AUG-1995
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/296,791
;      FILING DATE:  25-AUG-1994
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Trecartin, Richard F.
;      REGISTRATION NUMBER:  31,801
;      REFERENCE/DOCKET NUMBER:  FP-59941/RFT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1848 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
PCT-US95-10661A-6
```

Query Match 3.7%; Score 202.5; DB 5; Length 1848;
Best Local Similarity 18.5%; Pred. No. 8.7e-07;
Matches 177; Conservative 141; Mismatches 403; Indels 235; Gaps 34;

```
Qy     130 KQNGTSDGCSITFVRSTFVPASVGSQKVSPSTQSSQGKNADRST-LPKSVQEGNDSKCNA 188
           |:| | |  :||       |  :  |  :|| ||:  : |   | :|           |
Db     999 KRNQTVDTTNIT------TPNDI--QADAPSAQSNNEEIARVETPVPPP----------A 1040

Qy     189 PSGKNGAAEANTDSPMKDLQGPAQNYDVAANVSEDNTSVDVGALPEVPQITWHIEVNGAD 248
           |: ::  |    ::   :   ||      || :|     |     | |        |
Db    1041 PATESAIASEQPETRPAETAQPAMEETNTANSTETAPKSDTATQTENP--------NSES 1092

Qy     249 QPPSTPKLSEVVLKRNEDENGKTEETLVAEQCNLTKDPNPMSGK--ERDQVAEQCNLTKD 306
             |  |   :| | :    ||    ||:   :   | :| | :|:  : ||   : |   :
Db    1093 VPSET---TEKVAENPPQEN----ETVAKNEQEAT-EPTPQNGEVAKEDQPTVEANTQTN 1144

Qy     307 PKPVSGQKCE-----QICNEPCEEVVLKRSSKSKRKT----DKKLMKKQQHSKKRTAQAD 357
             | | |      :  :|| | | :  :   |  :    || :::|:: :|  | :
Db    1145 EATQSEGKTEETQTAETKSEPTESVTVSENQPEKTVSQSTEDKVVVEKEEKAKVETEETQ 1204

Qy     358 VSDAKLCRRKPKKVRLLSEIINANQVEDSRSDEVHRENAADPCEDDRSTIPVPMEVSMDI 417
           :     :  ||:              :   :||  :  |:  :  : | |  :  : :
Db    1205 KAPQVTSKEPPKQA------------EPAPEEVPTDTNAEEAQALQQTQPTTV-AAAET 1250

Qy     418 PVSNHTVGEDGLKSSKNKTKRKYSDVVDDGSSLM-----NWLNGKKKRTGSVHHTVAHPA 472
             |    |:  : |:       : || : ::            :|::|  |    :  :
Db    1251 TSPNSKPAEETQQPSEKTNAEPVTPVVSENTATQPTETEETAKVEKEKTQEVPQVASQES 1310

Qy     473 GNLSNKKVTPTASTQHDDENDTENGLDTNMHKTDVCQHVSEISTQRCSSKGKTAGLSKGK 532
                 | | |:   |   || | |         |   :  :    : |:||  ||
Db    1311 PKQEQPAAKPQAQTKPQAEPARENVLTTKNVGEPQPQAQPQTQSTAVPTTGETAANSKPA 1370

Qy     533 THSAASTKYGGESTRNGQNIHVLSAEDQCQMETENSVLSHSAKVSPAEHDIQIMSDLHEQ 592
             |  |   |  |  :|:  :: ::  | :|       || ||  :   :  |:: ||
Db    1371 AKPQAQAKPQTEPAR--ENVSTVNTKEP-QSQT-------SATVSTEQPAKETSSNV-EQ 1419

Qy     593 SLPKKKKKQKLEVTREKQTMIDDIPMDIVELLAKNQHERQLMTETDCSDINRIQSKTTAD 652
           |:         |  :     | |   :| : :|  : :  |  | |  |  :|
Db    1420 PAPENSINTGSATTMTETAEKSDKPQ--METVTENDRQPEANTVADNSVANNSESS---- 1473

Qy     653 DDCVIVAAKDGSDYASSVFDTNSQQKSLASQSTQKELQGHLALTTQE--------SPHPQ 704
                          :: |:::   ||  :   :     :|||        :| |:
```

```
Db   1474 ------------------ESKSRRRRSVSQPKETSAEETTVASTQETTVDNSVSTPKPR 1514

Qy    705 NFQSTQEQQTHLRMEEMVTIAASSPLFSHHDDQYIAEAPTEHWGRKDAKKLTWEQFKATT 764
          : :: :  ||:          :  |:          | |||:
Db   1515 SRRTRRSVQTN----------SYEPV----------ELPTEN------------------ 1536

Qy    765 RNSPAATCGAQFRPGIQAVDLTSTHVMGSSSNYASRQPVIAPLDRYAERAVNQVHARNFP 824
                  |:    :|           | :| |: ||           |:  : ::|
Db   1537 ---------AENAENVQ----------SGNNVANSQP----------ALRNLTSKNTN 1565

Qy    825 STIATMEASKLCDRRNAGQVVLYPKESMPATHLLRMMDPSTLASFPNYGT----SSRNQM 880
          : |:   |       | |: |       : |: ::          | |      |   |
Db   1566 AVISNAMAKAQFVALNVGKAV--------SQHISQL-------EMNNEGQYNVWISNTSM 1610

Qy    881 ESQLHNSQYAHNQYKGSTSTSYGSNLNGKIPLTFEDLSRHQLHDLHRPLRPHPRVGVLGS 940
              |  |:  ||:  :| |  : |     ::    :  ||  :   :|         |
Db   1611 -----NKNYSSEQYRRFSSKSTQTQLGWDQTIS----NNVQLGGVFTYVRNSNNFDKASS 1661

Qy    941 LLQKEIANWSENCGTQSGYKLGVSTGITSHQMNRKEH---------------FEALNSG 984
                |:       : : ||:  |    | | : :                 :| | |
Db   1662 KNTLAQVNFYSKYYADNHWYLGIDLGYGKFQSNLQTNNNAKFARHTAQIGLTAGKAFNLG 1721

Qy    985 MFSAKWNALQLGSVSSSADFLSARNSIAQSWTRGKGKMVHPLD---RFVRQDICIT 1037
            |: |       |  |:|||  |:: |           |:|:     | : |:  |
Db   1722 NFAVKPTVGVRYSYLSNADFALAQDRI----------KVNPISVKTAFAQVDLSYT 1767


RESULT   3
US-08-425-061-20
; Sequence 20, Application US/08425061
; Patent No. 5622829
;  GENERAL INFORMATION:
;    APPLICANT:  KING, Mary-Claire
;    APPLICANT:  FRIEDMAN, Lori
;    APPLICANT:  OSTERMEYER, Beth
;    APPLICANT:  ROWELL, Sarah
;    APPLICANT:  LYNCH, Eric
;    APPLICANT:  SZABO, Csilla
;    APPLICANT:  LEE, Ming
;    TITLE OF INVENTION:  GENETIC MARKERS FOR BREAST AND OVARIAN
;    TITLE OF INVENTION:  CANCER
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/425,061
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OSMAN, Richard A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  A-59563-3/DJB/RAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 494-8700
;      TELEFAX:  (415) 494-8771
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  900 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
```

```
US-08-425-061-20

  Query Match             3.3%;  Score 182;  DB 1;  Length 900;
  Best Local Similarity   18.4%;  Pred. No. 1.2e-05;
  Matches  200;  Conservative  142;  Mismatches  363;  Indels  382;  Gaps   49;

Qy      48 VDEPAQHQCEHFSIRGYVALL--QKKDPKFCSLSRIFHDQKKCDEHKASSSPFS--VAKF 103
           : ||   :|:|   :  :  |  ||| |  | | :   ::   |    |: ||  | :
Db      31 IKEPVSTKCDHIFCKFCMLKLLNQKKGPSQCPLCKNDITKRSLQE----STRFSQLVEEL 86

Qy     104 RRWDCSKCLD---------------------------------------------------- 113
             :  |:  ||
Db      87 LKIICAFQLDTGLEYANSYNFAKKENNSPEHLKDEVSIIQSMGYRNRAKRLLQSEPENPS 146

Qy     114 ------KLKTSDNGTAPRTLPAKQNGT----------SDGCSITFVRSTFVPASVGSQK 156
                 :: |: ||  |||  ||               ||    |  ::|:   ||| |:
Db     147 LQETSLSVQLSNLGTV-RTLRTKQRIQPQKTSVYIELGSDSSEDTVNKATY--CSVGDQE 203

Qy     157 VSPSTQSSQGKNADRSTLPKSVQEGNDSKCNAPSGKNGAAE-ANTDSPMKDLQGPAQNYD 215
           :  |           :|   :  | |  | | : ||    :    |: | |
Db     204 LLQIT----------------PQGTRDEISLDSAKKAACEFSETDVTNTEHHQPSNN-- 244

Qy     216 VAANVSEDNTSVDVGALPEVPQITWHIEVNGADQPPSTPKLSEVVLKRNEDENGKTEETL 275
                 : ||:                 |   |:: |             |   | :   |
Db     245 ------DLNTT----------------EKRAAERHP-------------EKYQGSSVSNL 269

Qy     276 VAEQCN----------------LTKDPNPMSGKERDQVAEQCNLTKDPKPVSGQK----- 314
             | |                 ||||      :   : || || :| |    |
Db     270 HVEPCGTNTHASSLQHENSSLLLTKD------RMNVEKAEFCNKSKQPGLARSQHNRWAG 323

Qy     315 CEQICNEPCEEVVLKRSSKSKRKTD---KKLMKKQQHSKKRTAQADVSDAKLCRRKPKK- 370
           :: ||:      :|:  :::| |     | :::: :|::           |   |:
Db     324 SKETCND-------RRTPSTEKKVDLNADPLCERKEWNKQKLP---------CSENPRDT 367

Qy     371 -----VRLLSEIINANQVEDSRSDEVHRENAADPCEDDRSTIPVPMEVSMDI-----PVS 420
                : | | |   |:   |||||:     :|   |  |     :   :|:       |
Db     368 EDVPWITLNSSIQKVNE-WFSRSDEL---LGSDDSHDGESESNAKVADVLDVLNEVDEYS 423

Qy     421 NHTVGEDGLKSSKN-----KTKRKYSDVVDDGSSLMNWLNGKKKRTGSVHHTVAH----- 470
             :  | | |  :      |::| :|  |:  |:: : : ||  |  :    ::|
Db     424 GSSEKIDLLASDPHEALICKSERVHSKSVE--SNIEDKIFGKTYRKKASLPNLSHVTENL 481

Qy     471 ---------------PAGNLSNKKVTPTASTQHDDENDTENGLDTNMHKTDVC-QHVSEI 514
                          |  |   :|  ||:    :|           : | |: |   |:
Db     482 IIGAFVTEPQIIQERPLTNKLKRKRRPTSGLHPED----------FIKKADLAVQKTPEM 531

Qy     515 ---STQRCSSKGKTAGLSKGKTHSAASTKYGGESTRNGQNIHVLSAEDQCQMETENSVLS 571
               | :    |:   :    |:|    |  |:| :| :| : : :     :| |::  :
Db     532 INQGTNQTEQNGQVMNI----TNSGHENKTKGDSIQNEKNPNPIES-----LEKESAFKT 582

Qy     572 HSAKVSPAEHDIQIMSDLHEQSLPKKKKKQKLEVTREKQTMIDDIPMDIVELLAKNQHER 631
             :  :| :  :::::  ::|    ||| : ::   ||           :  :||:
Db     583 KAEPISSSISNMELELNIHNSKAPKKNRLRRKSSTRH---------IHALELVVSRNLSP 633

Qy     632 QLMTE---TDCSDINRIQSK------TTADDDCVIVAAKDGSDYASSVFDTNSQQKSLAS 682
               ||     ||    |: |              :  ::  |: :  |       | |
Db     634 PNCTELQIDSCSSSEEIKKKKYNQMPVRHSRNLQLMEGKEPATGAKKSNKPNEQTSKRHD 693

Qy     683 QSTQKELQGHLALTTQESPHPQNF---QSTQEQQTHL--------RMEEMVTIAASSPLF 731
             |  ||:    ||       | :|    :| | :  :         : |:: |:  |:
Db     694 SDTFPELK----LTNA----PGSFTKCSNTSELKEFVNPSLPREEKEEKLETVKVSNN-- 743

Qy     732 SHHDDQYIAEAPTEHWGRKDAKKLTWEQFKATTRN--------SPAATCGAQFRPGIQAV 783
                 || |      ||   |: |:   | |:                |  | |     :::
Db     744 --------AEDP------KDL-MLSGERVLQTERSVESSSISLVPGTDYGTQ-----ESI 783

Qy     784 DLTSTHVMGSSSNYASRQPVIAPLDRYAERAVNQVHARNFPSTIATMEASKLCDRRNAGQ 843
           |     :|     |  :|          : |:|   |   |  :       ||  | ||  :
Db     784 SLLEVSTLGK----AKTEP---------NKCVSQCAAFENPKGL-IHGCSK--DNRNDTE 827

Qy     844 VVLYPKESMPATHLLRMMDPSTLASFPNYGTSSRNQMESQLHNSQYAHNQYKGSTSTSYG 903
             ||                 |    |:   :  :||    ::||  | :| |   |:
```

```
Db    828 GFKYP----------------LGHEVNHSRETSIEMEESELDAQYLQNTFKVSKRQSFA 870

Qy    904 SNLNGKIPLTFEDLSRHQLHDLHRPLRPHPRVGVLGSLLQKEIANWSENCGTQSGYKLGV 963

Db    871 -------------------------PFSNPG----------NAEEECATFSAHSGSL 892

Qy    964 STGITSH 970

Db    893 KTKSKSH 899


RESULT   4
US-08-825-886-20
; Sequence 20, Application US/08825886
; Patent No. 5821328
;  GENERAL INFORMATION:
;    APPLICANT:  KING, Mary-Claire
;    APPLICANT:  FRIEDMAN, Lori
;    APPLICANT:  OSTERMEYER, Beth
;    APPLICANT:  ROWELL, Sarah
;    APPLICANT:  LYNCH, Eric
;    APPLICANT:  SZABO, Csilla
;    APPLICANT:  LEE, Ming
;    TITLE OF INVENTION:  GENETIC MARKERS FOR BREAST AND OVARIAN
;    TITLE OF INVENTION:  CANCER
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/825,886
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/425,061
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OSMAN, Richard A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  A-59563-3/DJB/RAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 494-8700
;      TELEFAX:  (415) 494-8771
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  900 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-825-886-20


  Query Match             3.3%;  Score 182;  DB 2;  Length 900;
  Best Local Similarity   18.4%;  Pred. No. 1.2e-05;
  Matches  200;  Conservative  142;  Mismatches  363;  Indels  382;  Gaps   49;

Qy     48 VDEPAQHQCEHFSIRGYVALL--QKKDPKFCSLSRIFHDQKKCDEHKASSSPFS--VAKF 103

Db     31 IKEPVSTKCDHIFCKFCMLKLLNQKKGPSQCPLCKNDITKRSLQE----STRFSQLVEEL 86

Qy    104 RRWDCSKCLD---------------------------------------------------- 113


Db     87 LKIICAFQLDTGLEYANSYNFAKKENNSPEHLKDEVSIIQSMGYRNRAKRLLQSEPENPS 146

Qy    114 ------KLKTSDNGTAPRTLPAKQNGT-----------SDGCSITFVRSTFVPASVGSQK 156

Db    147 LQETSLSVQLSNLGTV-RTLRTKQRIQPQKTSVYIELGSDSSEDTVNKATY--CSVGDQE 203

Qy    157 VSPSTQSSQGKNADRSTLPKSVQEGNDSKCNAPSGKNGAAE-ANTDSPMKDLQGPAQNYD 215

Db    204 LLQIT----------------PQGTRDEISLDSAKKAACEFSETDVTNTEHHQPSNN-- 244

Qy    216 VAANVSEDNTSVDVGALPEVPQITWHIEVNGADQPPSTPKLSEVVLKRNEDENGKTEETL 275

Db    245 ------DLNTT----------------EKRAAERHP------------EKYQGSSVSNL 269

Qy    276 VAEQCN----------------LTKDPNPMSGKERDQVAEQCNLTKDPKPVSGQK----- 314

Db    270 HVEPCGTNTHASSLQHENSSLLLTKD------RMNVEKAEFCNKSKQPGLARSQHNRWAG 323

Qy    315 CEQICNEPCEEVVLKRSSKSKRKTD---KKLMKKQQHSKKRTAQADVSDAKLCRRKPKK- 370

Db    324 SKETCND-------RRTPSTEKKVDLNADPLCERKEWNKQKLP---------CSENPRDT 367

Qy    371 -----VRLLSEIINANQVEDSRSDEVHRENAADPCEDDRSTIPVPMEVSMDI-----PVS 420

Db    368 EDVPWITLNSSIQKVNE-WFSRSDEL---LGSDDSHDGESESNAKVADVLDVLNEVDEYS 423

Qy    421 NHTVGEDGLKSSKN-----KTKRKYSDVVDDGSSLMNWLNGKKKRTGSVHHTVAH----- 470

Db    424 GSSEKIDLLASDPHEALICKSERVHSKSVE--SNIEDKIFGKTYRKKASLPNLSHVTENL 481

Qy    471 ----------------PAGNLSNKKVTPTASTQHDDENDTENGLDTNMHKTDVC-QHVSEI 514

Db    482 IIGAFVTEPQIIQERPLTNKLKRKRRPTSGLHPED----------FIKKADLAVQKTPEM 531

Qy    515 ---STQRCSSKGKTAGLSKGKTHSAASTKYGGESTRNGQNIHVLSAEDQCQMETENSVLS 571

Db    532 INQGTNQTEQNGQVMNI----TNSGHENKTKGDSIQNEKNPNPIES-----LEKESAFKT 582

Qy    572 HSAKVSPAEHDIQIMSDLHEQSLPKKKKKQKLEVTREKQTMIDDIPMDIVELLAKNQHER 631

Db    583 KAEPISSSISNMELELNIHNSKAPKKNRLRRKSSTRH---------IHALELVVSRNLSP 633

Qy    632 QLMTE---TDCSDINRIQSK------TTADDDCVIVAAKDGSDYASSVFDTNSQQKSLAS 682

Db    634 PNCTELQIDSCSSSEEIKKKKYNQMPVRHSRNLQLMEGKEPATGAKKSNKPNEQTSKRHD 693

Qy    683 QSTQKELQGHLALTTQESPHPQNF---QSTQEQQTHL--------RMEEMVTIAASSPLF 731

Db    694 SDTFPELK----LTNA----PGSFTKCSNTSELKEFVNPSLPREEKEEKLETVKVSNN-- 743

Qy    732 SHHDDQYIAEAPTEHWGRKDAKKLTWEQFKATTRN--------SPAATCGAQFRPGIQAV 783

Db    744 --------AEDP------KDL-MLSGERVLQTERSVESSSISLVPGTDYGTQ-----ESI 783

Qy    784 DLTSTHVMGSSSNYASRQPVIAPLDRYAERAVNQVHARNFPSTIATMEASKLCDRRNAGQ 843

Db    784 SLLEVSTLGK----AKTEP---------NKCVSQCAAFENPKGL-IHGCSK--DNRNDTE 827

Qy    844 VVLYPKESMPATHLLRMMDPSTLASFPNYGTSSRNQMESQLHNSQYAHNQYKGSTSTSYG 903

Db    828 GFKYP----------------LGHEVNHSRETSIEMEESELDAQYLQNTFKVSKRQSFA 870

Qy    904 SNLNGKIPLTFEDLSRHQLHDLHRPLRPHPRVGVLGSLLQKEIANWSENCGTQSGYKLGV 963

Db    871 -------------------------PFSNPG----------NAEEECATFSAHSGSL 892

Qy    964 STGITSH 970

Db    893 KTKSKSH 899


RESULT   5
US-08-603-753D-2
```

; Sequence 2, Application US/08603753D
; Patent No. 5891857
; GENERAL INFORMATION:
; APPLICANT: HOLT, JEFFREY T.
; APPLICANT: JENSEN, ROY A.
; APPLICANT: PAGE, DAVID L.
; APPLICANT: KING, MARY-CLAIRE
; APPLICANT: SZABO, CSILLA I.
; APPLICANT: JETTON, THOMAS L.
; APPLICANT: ROBINSON-BENION, CHERYL L.
; APPLICANT: THOMPSON, MARILYN E.
; TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2
; TITLE OF INVENTION: PROTEINS AND SCREENING AND THERAPEUTIC METHODS BASED ON
; TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2 PROTEINS.
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARLES A. TAYLOR, JR.
; STREET: SUITE 1401, UNIVERSITY TOWER, 3100 TOWER
; STREET: BOULEVARD
; CITY: DURHAM
; STATE: NORTH CAROLINA
; COUNTRY: USA
; ZIP: 27707
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 800 kB storage
; COMPUTER: IBM PC/XT/AT compatible
; OPERATING SYSTEM: Windows 3.1
; SOFTWARE: WORD PERFECT 6.1 and ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/603,753D
; FILING DATE: 20 FEB 1996
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.S. 08/373,799
; FILING DATE: 17 JAN 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: ARLES A. TAYLOR, JR.
; REGISTRATION NUMBER: 39,395
; REFERENCE/DOCKET NUMBER: 1242/2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919) 493-8000
; TELEFAX: (919) 419-0383
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1863
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHETICAL: no
; ANTI-SENSE: no
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; INDIVIDUAL ISOLATE:
; DEVELOPMENTAL STAGE: adult
; TISSUE TYPE: female breast
; CELL TYPE: ductal carcinoma in situ, invasive
; CELL TYPE: breast cancer and normal breast tissue
; CELL LINE: not derived from a cell line
; ORGANELLE: no
; IMMEDIATE SOURCE:
; LIBRARY: cDNA library derived from human
; CLONE: obtained using published sequence
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: unknown
; MAP POSITION: unknown
; UNITS: unknown
; FEATURE:
; NAME/KEY: BRCA1 protein
; LOCATION: 1 to 1863
; IDENTIFICATION METHOD: observation of mRNA and
; IDENTIFICATION METHOD: antisense inhibition of BRCA1 gene
; OTHER INFORMATION: BRCA1 protein has a negative
; OTHER INFORMATION: regulatory effect on growth of human mammary cells.
; PUBLICATION INFORMATION:
; AUTHORS: Miki, Y., et. al.
; TITLE: A strong candidate gene for the breast and
; TITLE: ovarian cancer susceptibility gene BRCA1.
; JOURNAL: Science
; VOLUME: 266
; PAGES: 66-71
; DATE: 1994
; RELEVANT RESIDUES IN SEQ ID NO: 2: granin box
; RELEVANT RESIDUES IN SEQ ID NO: domain at amino acids 1214-1223
US-08-603-753D-2

Query Match 3.3%; Score 180; DB 2; Length 1863;
Best Local Similarity 17.9%; Pred. No. 5.2e-05;
Matches 221; Conservative 174; Mismatches 387; Indels 452; Gaps 62;

```
Qy      48 VDEPAQHQCEHFSIRGYVALL--QKKDPKFCSLSRIFHDQKKCDEHKASSSPFS--VAKF 103
           : ||   :|:|   :  :  |  ||| |  | | :    ::   |    |: ||  | :
Db      31 IKEPVSTKCDHIFCKFCMLKLLNQKKGPSQCPLCKNDITKRSLQE---STRFSQLVEEL 86

Qy     104 RRWDCSKCLD---------------------------------------------------- 113
            :  |:  ||
Db      87 LKIICAFQLDTGLEYANSYNFAKKENNSPEHLKDEVSIIQSMGYRNRAKRLLQSEPENPS 146

Qy     114 ------KLKTSDNGTAPRTLPAKQNGT----------SDGCSITFVRSTFVPASVGSQK 156
                 :: |: ||  |||  ||              ||    |  ::|:   ||| |:
Db     147 LQETSLSVQLSNLGTV-RTLRTKQRIQPQKTSVYIELGSDSSEDTVNKATY--CSVGDQE 203

Qy     157 VSPSTQSSQGKNADRSTLPKSVQEGNDSKCNAPSGKNGAAE-ANTDSPMKDLQGPAQNYD 215
           :  |                  :|   : :  | |  | | : ||    :   |: |
Db     204 LLQIT----------------PQGTRDEISLDSAKKAACEFSETDVTNTEHHQPSNN-- 244

Qy     216 VAANVSEDNTSVDVGALPEVPQ-------ITWHIEVNGADQPPSTPKLSEVVLKRNEDEN 268
                 |  : :  |    |:         |:|  | :   |: :     |   :|
Db     245 -------DLNTTEKRAAERHPEKYQGSSVSNLHVEPCGTNTHASSLQHENSSLLLTKDRM 297

Qy     269 GKTEETLVAEQCNLTKDPNPMSGKERDQ------VAEQCNLTKDPKPVSGQK-----CEQ 317
              :   || || :| |    |  | |        | ||   | :  | :|      :
Db     298 NVEK----AEFCNKSKQP----GLARSQHNRWAGSKETCN---DRRTPSTEKKVDLNADP 346

Qy     318 ICNE--------PCEE----------VVLKRSSKSKR----KTDKKLMKKQQHSKKRTAQ 355
           :|          || |          : |  | :       ::|: |      |  :  :
Db     347 LCERKEWNKQKLPCSENPRDTEDVPWITLNSSIQKVNEWFSRSDELLGSDDSHDGESESN 406

Qy     356 ADVSDAKLCRRK-------PKKVRLLSEIINANQVEDSRSDEVHRENAADPCED------ 402
           | |:|      :        :|: ||:   : ::    :|| || ::     ||
Db     407 AKVADVLDVLNEVDEYSGSSEKIDLLAS--DPHEALICKSDRVHSKSVESDIEDKIFGKT 464

Qy     403 --DRSTIPVPMEVSMDI--------------PVSNH-------TVG---EDGLKSSKNK 435
              ::::|    |: ::              |::|        | |   || :| : :
Db     465 YRKKASLPNLSHVTENLIIGAFVSEPQIIQERPLTNKLKRKRRPTSGLHPEDFIKKA-DL 523

Qy     436 TKRKYSDVVDDGSS-------LMNWLN-GKKKRT-----------------GSVHHTV 468
             :|  :::: |::       :||  | | : :|                    |   |
Db     524 AVQKTPEMINQGTNQTEQNGQVMNITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTK 583

Qy     469 AHP-AGNLSNKKVTPTASTQHDDENDTENGL-----DTNMHKTDVC-------QHVSEIS 515
           | | : ::||:       :  |: :   :| |        ::|  ::          : :|:
Db     584 AEPISSSISNEL---ELNIMHNSKAPKKNRLRRKSSTRHIHALELVVSRNLSPPNCTELQ 640

Qy     516 TQRCSSKGKTAGLSKGKTHSAASTKYGGESTRNGQNIHVLSAED------QCQMETENSV 569
              |||           :      ||     |: :|: ::  ::      :     | :
Db     641 IDSCSS----------SEEIKKKKYNQMPVRHSRNLQLMEGKEPATGAKKSNKPNEQTS 689

Qy     570 LSHSAKVSPAEHDIQI-----------MSDLHE---QSLPKKKKKQKLEVTREKQTMID 614
             | :   |   ::::             |:| |    |||:::|::|||  :      |
Db     690 KRHDSDTFP---ELKLTNAPGSFTKCSNTSELKEFVNPSLPREEKEEKLETVKVSNNAED 746

Qy     615 DIPMDIVELLAKNQHERQLMTETDCSDINRIQSKTTADDDCVIVAAKDGSDY-------- 666
             | |::        || | ||        ::| :        ::   |:||
```

```
Db    747 --PKDLM-----LSGERVLQTE------RSVESSS---------ISLVPGTDYGTQESISL 785

Qy    667 --ASSVFDTNSQQKSLASQSTQKELQGHL--------AITTQESPHPQNFQSTQEQQTHL 716

Db    786 LEVSTLGKAKTEPNKCVSQCAAFENPKGLIHGCSKDNRMDTEGFKYPLGHEVNHSRETSI 845

Qy    717 RMEEMVTIAASSPLFSHHDDQYIAE----------AP--------------TEHWG--RK 750

Db    846 EMEE-----------SELDAQYLQNTFKVSKRQSFAPFSNPGNAEEECATFSAHSGSLKK 894

Qy    751 DAKKLTWE-------QFKATTRNSPAAT---------CQAQFRPGIQAVDLTSTHVMGSS 794

Db    895 QSPKVTFECEQKEENQGKNESNIKPVQTVNITAGFPVVGQKDKP----VDNAKCSIKGGS 950

Qy    795 SNYASRQ----------------------PVIAPLDRYAE-RAVNQVHARNFPSTIATM 830

Db    951 RFCLSSQFRGNETGLITPNKHGLLQNPYRIPPLFPIKSIVKTKCKKNLLEENF------- 1003

Qy    831 EASKLCDRRNAGQVVLYPKESMPATHLLRMMDPSTLASIPNYGTSSRNQMESQLHNSQYA 890

Db   1004 EEHSMSPEREMG------NENIPST---------------VSTISRNNIRENVFKEASS 1041

Qy    891 HN-QYKGSTSTSYGSNLNGKIPLTFE----DLSRHQLHDLHRPLRPHPRVGVLGSLLQKE 945

Db   1042 SNINEVGSSTNEVGSSIN-EIGSSDENIQAELGRNRGPILNAMLR-------LG-VLQPE 1092

Qy    946 IANWSENCGTQSGYKLGVSTGITSHQMNRKEHFE 979

Db   1093 V-----------YKQSLPGSNCKHPEIKKQEYE 1114


RESULT   6
US-09-099-753-2
; Sequence 2, Application US/09099753
; Patent No. 6149903
;  GENERAL INFORMATION:
;    APPLICANT:  HOLT, JEFFREY T.
;    APPLICANT:  JENSEN, ROY A.
;    APPLICANT:  PAGE, DAVID L.
;    APPLICANT:  KING, MARY-CLAIRE
;    APPLICANT:  SZABO, CSILLA I.
;    APPLICANT:  JETTON, THOMAS L.
;    APPLICANT:  ROBINSON-BENION, CHERYL L.
;    APPLICANT:  THOMPSON, MARILYN E.
;    TITLE OF INVENTION:  CHARACTERIZED BRCA1 AND BRCA2
;    TITLE OF INVENTION:  PROTEINS AND SCREENING AND THERAPEUTIC METHODS BASED ON
;    TITLE OF INVENTION:  CHARACTERIZED BRCA1 AND BRCA2 PROTEINS.
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ARLES A. TAYLOR, JR.
;      STREET:  SUITE 1401, UNIVERSITY TOWER, 3100 TOWER
;      STREET:  BOULEVARD
;      CITY:  DURHAM
;      STATE:  NORTH CAROLINA
;      COUNTRY:  USA
;      ZIP:  27707
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 800 kB storage
;      COMPUTER:  IBM PC/XT/AT compatible
;      OPERATING SYSTEM:  Windows 3.1
;      SOFTWARE:  WORD PERFECT 6.1 and ASCII
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/099,753
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/603,753
;      FILING DATE:  20 FEB 1996
;      APPLICATION NUMBER:  U.S. 08/373,799
;      FILING DATE:  17 JAN 1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ARLES A. TAYLOR, JR.
;      REGISTRATION NUMBER:  39,395
;      REFERENCE/DOCKET NUMBER:  1242/2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (919) 493-8000
;      TELEFAX:  (919) 419-0383
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1863
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  no
;    ANTI-SENSE:  no
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;      INDIVIDUAL ISOLATE:
;      DEVELOPMENTAL STAGE:  adult
;      TISSUE TYPE:  female breast
;      CELL TYPE:  ductal carcinoma in situ, invasive
;      CELL TYPE:  breast cancer and normal breast tissue
;      CELL LINE:  not derived from a cell line
;      ORGANELLE:  no
;    IMMEDIATE SOURCE:
;      LIBRARY:  cDNA library derived from human
;      CLONE:  obtained using published sequence
;    POSITION IN GENOME:
;      CHROMOSOME/SEGMENT:  unknown
;      MAP POSITION:  unknown
;      UNITS:  unknown
;    FEATURE:
;      NAME/KEY:  BRCA1 protein
;      LOCATION:  1 to 1863
;      IDENTIFICATION METHOD:  observation of mRNA and
;      IDENTIFICATION METHOD:  antisense inhibition of BRCA1 gene
;      OTHER INFORMATION:  BRCA1 protein has a negative
;      OTHER INFORMATION:  regulatory effect on growth of human mammary cells.
;    PUBLICATION INFORMATION:
;      AUTHORS:  Miki, Y., et. al.
;      TITLE:  A strong candidate gene for the breast and
;      TITLE:  ovarian cancer susceptibility gene BRCA1.
;      JOURNAL:  Science
;      VOLUME:  266
;      PAGES:  66-71
;      DATE:  1994
;      RELEVANT RESIDUES IN SEQ ID NO:  2:  granin box
;      RELEVANT RESIDUES IN SEQ ID NO:  domain at amino acids 1214-1223
US-09-099-753-2


  Query Match             3.3%;  Score 180;  DB 4;  Length 1863;
  Best Local Similarity   17.9%;  Pred. No. 5.2e-05;
  Matches  221;  Conservative  174;  Mismatches  387;  Indels  452;  Gaps   62;

Qy     48 VDEPAQHQCEHFSIRGYVALL--QKKDPKFCSLSRIFHDQKKCDEHKASSSPFS--VAKF 103

Db     31 IKEPVSTKCDHIFCKFCMLKLLNQKKGPSQCPLCKNDITKRSLQE----STRFSQLVEEL 86

Qy    104 RRWDCSKCLD---------------------------------------------------- 113

Db     87 LKIICAFQLDTGLEYANSYNFAKKENNSPEHLKDEVSIIQSMGYRNRAKRLLQSEPENPS 146

Qy    114 ------KLKTSDNGTAPRTLPAKQNGT----------SDGCSITFVRSTFVPASVGSQK 156

Db    147 LQETSLSVQLSNLGTV-RTLRTKQRIQPQKTSVYIELGSDSSEDTVNKATY--CSVGDQE 203

Qy    157 VSPSTQSSQGKNADRSTLPKSVQEGNDSKCNAPSGKNGAAE-ANTDSPMKDLQGPAQNYD 215

Db    204 LLQIT----------------PQGTRDEISLDSAKKAACEFSETDVTNTEHHQPSNN-- 244

Qy    216 VAANVSEDNTSVDVGALPEVPQ-------ITWHIEVNGADQPPSTPKLSEVVLKRNEDEN 268

Db    245 -------DLNTTEKRAAERHPEKYQGSSVSNLHVEPCGTNTHASSLQHENSSLLLTKDRM 297
```

```
Qy    269 GKTEETLVAEQCNLTKDPNPMSGKERDQ------VAEQCNLTKDPKPVSGQK-----CEQ 317
Db    298 NVEK----AEFCNKSKQP----GLARSQHNRWAGSKETCN---DRRTPSTEKKVDLNADP 346

Qy    318 ICNE--------PCEE----------VVLKRSSKSKR----KTDKKLMKKQQHSKKRTAQ 355
Db    347 LCERKEWNKQKLPCSENPRDTEDVPWITLNSSIQKVNEWFSRSDELLGSDDSHDGESESN 406

Qy    356 ADVSDAKLCRRK-------PKKVRLLSEIINANQVEDSRSDEVHRENAADPCED------ 402
Db    407 AKVADVLDVLNEVDEYSGSSEKIDLLAS--DPHEALICKSDRVHSKSVESDIEDKIFGKT 464

Qy    403 --DRSTIPVPMEVSMDI--------------PVSNH-------TVG---EDGLKSSKNK 435
Db    465 YRKKASLPNLSHVTENLIIGAFVSEPQIIQERPLTNKLKRKRRPTSGLHPEDFIKKA-DL 523

Qy    436 TKRKYSDVVDDGSS-------LMNWLN-GKKKRT------------------GSVHHTV 468
Db    524 AVQKTPEMINQGTNQTEQNGQVMNITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTK 583

Qy    469 AHP-AGNLSNKKVTPTASTQHDDENDTENGL-----DTNMHKIDVC-------QHVSEIS 515
Db    584 AEPISSSISNEL---ELNIMHNSKAPKKNRLRRKSSTRHIHALELVVSRNLSPPNCTELQ 640

Qy    516 TQRCSSKGKTAGLSKGKTHSAASTKYGGESTRNGQNIHVLSAED------QCQMETENSV 569
Db    641 IDSCSS-----------SEEIKKKKYNQMPVRHSRNLQLMEGKEPATGAKKSNKPNEQTS 689

Qy    570 LSHSAKVSPAEHDIQI------------MSDLHE---QSLPKKKKKQKLEVTREKQTMID 614
Db    690 KRHDSDTFP---ELKLTNAPGSFTKCSNTSELKEFVNPSLPREEKEEKLETVKVSNNAED 746

Qy    615 DIPMDIVELLAKNQHERQLMTETDCSDINRIQSKTTADDDCVIVAAKDGSDY-------- 666
Db    747 --PKDLM-----LSGERVLQTE------RSVESSS--------ISLVPGTDYGTQESISL 785

Qy    667 --ASSVFDTNSQQKSLASQSTQKELQGHL--------ALTTQESPHPQNFQSTQEQQTHL 716
Db    786 LEVSTLGKAKTEPNKCVSQCAAFENPKGLIHGCSKDNRNDTEGFKYPLGHEVNHSRETSI 845

Qy    717 RMEEMVTIAASSPLFSHHDDQYIAE----------AP--------------TEHWG--RK 750
Db    846 EMEE-----------SELDAQYLQNTFKVSKRQSFAPFSNPGNAEEECATFSAHSGSLKK 894

Qy    751 DAKKLTWE-------QFKATTRNSPAAT---------CGAQFRPGIQAVDLTSTHVMGSS 794
Db    895 QSPKVTFECEQKEENQGKNESNIKPVQTVNITAGFPVVGQKDKP----VDNAKCSIKGGS 950

Qy    795 SNYASRQ----------------------PVIAPLDRYAE-RAVNQVHARNFPSTIATM 830
Db    951 RFCLSSQFRGNETGLITPNKHGLLQNPYRIPPLFPIKSFVKTKCKKNLLEENF------- 1003

Qy    831 EASKLCDRRNAGQVVLYPKESMPATHLLRMMDPSTLASFPNYGTSSRNQMESQLHNSQYA 890
Db   1004 EEHSMSPEREMG------NENIPST---------------VSTISRNNIRENVFKEASS 1041

Qy    891 HN-QYKGSTSTSYGSNLNGKIPLTFE----DLSRHQLHDLHRPLRPHPRVGVLGSLLQKE 945
Db   1042 SNINEVGSSTNEVGSSIN-EIGSSDENIQAELGRNRGPKLNAMLR-------LG-VLQPE 1092

Qy    946 IANWSENCGTQSGYKLGVSTGITSHQMNRKEHFE 979
Db   1093 V------------YKQSLPGSNCKHPEIKKQEYE 1114
```

```
RESULT   7
US-08-986-106-2
; Sequence 2, Application US/08986106
; Patent No. 6177410
;  GENERAL INFORMATION:
;    APPLICANT:  HOLT, JEFFREY T.
;    APPLICANT:  JENSEN, ROY A.
;    APPLICANT:  KING, MARY-CLAIRE
;    APPLICANT:  STEINER, MITCHELL S.
;    APPLICANT:  ROBINSON-BENION, CHERYL L.
;    APPLICANT:  THOMPSON, MARILYN E.
;    TITLE OF INVENTION:  THERAPEUTIC METHODS FOR
;    TITLE OF INVENTION:  PROSTATE CANCER
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ARLES A. TAYLOR, JR.
;      STREET:  SUITE 1401, UNIVERSITY TOWER, 3100 TOWER
;      STREET:  BOULEVARD
;      CITY:  DURHAM
;      STATE:  NORTH CAROLINA
;      COUNTRY:  USA
;      ZIP:  27707
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.44MB storage
;      COMPUTER:  IBM PC/XT/AT compatible
;      OPERATING SYSTEM:  Windows 3.1
;      SOFTWARE:  WORD PERFECT 6.1 and ASCII
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/986,106
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/603,753
;      FILING DATE:  20 FEB 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ARLES A. TAYLOR, JR.
;      REGISTRATION NUMBER:  39,395
;      REFERENCE/DOCKET NUMBER:  1242/3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (919) 493-8000
;      TELEFAX:  (919) 419-0383
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1863
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  unknown
;    PUBLICATION INFORMATION:
;      AUTHORS:  Miki, Y., et. al.
;      TITLE:  A strong candidate gene for the breast and
;      TITLE:  ovarian cancer susceptibility gene
;      TITLE:  BRCA1.
;      JOURNAL:  Science
;      VOLUME:  266
;      PAGES:  66-71
;      DATE:  1994
;      RELEVANT RESIDUES IN SEQ ID NO:  2:  granin box domain
;      RELEVANT RESIDUES IN SEQ ID NO:  at amino acids 1214-1223
US-08-986-106-2


  Query Match             3.3%;  Score 180;  DB 4;  Length 1863;
  Best Local Similarity   17.9%;  Pred. No. 5.2e-05;
  Matches  221;  Conservative  174;  Mismatches  387;  Indels  452;  Gaps   62;

Qy     48 VDEPAQHQCEHFSIRGYVALL--QKKDPKFCSLSRIFHDQKKCDEHKASSSPFS--VAKF 103
Db     31 IKEPVSTKCDHIFCKFCMLKLLNQKKGPSQCPLCKNDITKRSLQE----STRFSQLVEEL 86

Qy    104 RRWDCSKCLD---------------------------------------------------- 113
Db     87 LKIICAFQLDTGLEYANSYNFAKKENNSPEHLKDEVSIIQSMGYRNRAKRLLQSEPENPS 146

Qy    114 ------KLKTSDNGTAPRTLPAKQNGT----------SDGCSITFVRSTFVPASVGSQK 156
Db    147 LQETSLSVQLSNLGTV-RTLRTKQRIQPQKTSVYIELGSDSSEDTVNKATY--CSVGDQE 203

Qy    157 VSPSTQSSQGKNADRSTLPKSVQEGNDSKCNAPSGKNGAAE-ANTDSPMKDLQGPAQNYD 215
Db    204 LLQIT----------------PQGTRDEISLDSAKKAACEFSETDVTNTEHHQPSNN-- 244
```

```
Qy    216 VAANVSEDNTSVDVGALPEVPQ-------ITWHIEVNGADQPPSTPKLSEVVLKRNEDEN 268
Db    245 -------DLNTTEKRAAERHPEKYQGSSVSNLHVEPCGTNTHASSLQHENSSLLLTKDRM 297

Qy    269 GKTEETLVAEQCNLTKDPNPMSGKERDQ------VAEQCNLTKDPKPVSGQK-----CEQ 317
Db    298 NVEK----AEFCNKSKQP----GLARSQHNRWAGSKETCN---DRRTPSTEKKVDLNADP 346

Qy    318 ICNE--------PCEE----------VVLKRSSKSKR----KTDKKLMKKQQHSKKRTAQ 355
Db    347 LCERKEWNKQKLPCSENPRDTEDVPWITLNSSIQKVNEWFSRSDELLGSDDSHDGESESN 406

Qy    356 ADVSDAKLCRRK-------PKKVRLLSEIINANQVEDSRSDEVHRENAADPCED------ 402
Db    407 AKVADVLDVLNEVDEYSGSSEKIDLLAS--DPHEALICKSDRVHSKSVESDIEDKIFGKT 464

Qy    403 --DRSTIPVPMEVSMDI---------------PVSNH-------TVG---EDGLKSSKNK 435
Db    465 YRKKASLPNLSHVTENLIIGAFVSEPQIIQERPLTNKLKRKRRPTSGLHPEDFIKKA-DL 523

Qy    436 TKRKYSDVVDDGSS-------LMNWLN-GKKKRT-------------------GSVHHTV 468
Db    524 AVQKTPEMINQGTNQTEQNGQVMNITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTK 583

Qy    469 AHP-AGNLSNKKVTPTASTQHDDENDTENGL-----DTNMHKTDVC-------QHVSEIS 515
Db    584 AEPISSSISNEL---ELNIMHNSKAPKKNRLRRKSSTRHIHALELVVSRNLSPPNCTELQ 640

Qy    516 TQRCSSKGKTAGLSKGKTHSAASTKYGGESTRNGQNIHVLSAED------QCQMETENSV 569
Db    641 IDSCSS-----------SEEIKKKKYNQMPVRHSRNLQLMEGKEPATGAKKSNKPNEQTS 689

Qy    570 LSHSAKVSPAEHDIQI-----------MSDLHE---QSLPKKKKKQKLEVTREKQTMID 614
Db    690 KRHDSDTFP---ELKLTNAPGSFTKCSNTSELKEFVNPSLPREEKEEKLETVKVSNNAED 746

Qy    615 DIPMDIVELLAKNQHERQLMTETDCSDINRIQSKTTADDDCVIVAAKDGSDY-------- 666
Db    747 --PKDLM-----LSGERVLQTE------RSVESSS---------ISLVPGTDYGTQESISL 785

Qy    667 --ASSVFDTNSQQKSLASQSTQKELQGHL--------ALTTQESPHPQNFQSTQEQQTHL 716
Db    786 LEVSTLGKAKTEPNKCVSQCAAFENPKGLIHGCSKDNRNDTEGFKYPLGHEVNHSRETSI 845

Qy    717 RMEEMVTIAASSPLFSHHDDQYIAE----------AP--------------TEHWG--RK 750
Db    846 EMEE-----------SELDAQYLQNTFKVSKRQSFAPFSNPGNAEEECATFSAHSGSLKK 894

Qy    751 DAKKLTWE-------QFKATTRNSPAAT---------CGAQFRPGIQAVDLTSTHVMGSS 794
Db    895 QSPKVTFECEQKEENQGKNESNIKPVQTVNITAGFPVVGQKDKP----VDNAKCSIKGGS 950

Qy    795 SNYASRQ----------------------PVIAPLDRYAE-RAVNQVHARNFPSTIATM 830
Db    951 RFCLSSQFRGNETGLITPNKHGLLQNPYRIPPLFPIKSFVKTKCKKNLLEENF------- 1003

Qy    831 EASKLCDRRNAGQVVLYPKESMPATHLLRMMDPSTLASFPNYGTSSRNQMESQLHNSQYA 890
Db   1004 EEHSMSPEREMG------NENIPST---------------VSTISRNNIRENVFKEASS 1041

Qy    891 HN-QYKGSTSTSYGSNLNGKIPLTFE----DLSRHQLHDLHRPLRPHPRVGVLGSLLQKE 945
Db   1042 SNINEVGSSTNEVGSSIN-EIGSSDENIQAELGRNRGPKLNAMLR-------LG-VLQPE 1092

Qy    946 IANWSENCGTQSGYKLGVSTGITSHQMNRKEHFE 979
Db   1093 V------------YKQSLPGSNCKHPEIKKQEYE 1114


RESULT   8
US-09-007-678B-49
; Sequence 49, Application US/09007678B
; Patent No. 6342483
; GENERAL INFORMATION:
;  APPLICANT: HOLT, JEFFREY T.
;  APPLICANT:  JENSEN, ROY A.
;  APPLICANT:  PAGE, DAVID L.
;  APPLICANT:  OBERMILLER, PATRICE S.
;  APPLICANT:  ROBINSON-BENION, CHERYL L.
;  APPLICANT:  THOMPSON, MARILYN E.
;  TITLE OF INVENTION: METHOD FOR DETECTION AND TREATMENT OF BREAST CANCER
;  FILE REFERENCE: Attorney Docket No. 6342483 1242-1-2-2
;  CURRENT APPLICATION NUMBER: US/09/007,678B
;  CURRENT FILING DATE:  1998-01-15
;  PRIOR APPLICATION NUMBER: 08/373,799
;  PRIOR FILING DATE: 1995-01-17
;  PRIOR APPLICATION NUMBER: 08/182,961
;  PRIOR FILING DATE: 1994-01-14
;  NUMBER OF SEQ ID NOS: 61
;  SOFTWARE: MicroSoft WordPad
; SEQ ID NO 49
;   LENGTH: 1863
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1472)
;   OTHER INFORMATION: Xaa=any amino acid
US-09-007-678B-49


  Query Match             3.3%;  Score 180;  DB 4;  Length 1863;
  Best Local Similarity   17.9%;  Pred. No. 5.2e-05;
  Matches  221;  Conservative  174;  Mismatches  387;  Indels  452;  Gaps   62;

Qy     48 VDEPAQHQCEHFSIRGYVALL--QKKDPKFCSLSRIFHDQKKCDEHKASSSPFS--VAKF 103
Db     31 IKEPVSTKCDHIFCKFCMLKLLNQKKGPSQCPLCKNDITKRSLQE----STRFSQLVEEL 86

Qy    104 RRWDCSKCLD---------------------------------------------------- 113
Db     87 LKIICAFQLDTGLEYANSYNFAKKENNSPEHLKDEVSIIQSMGYRNRAKRLLQSEPENPS 146

Qy    114 ------KLKTSDNGTAPRTLPAKQNGT----------SDGCSITFVRSTFVPASVGSQK 156
Db    147 LQETSLSVQLSNLGTV-RTLRTKQRIQPQKTSVYIELGSDSSEDTVNKATY--CSVGDQE 203

Qy    157 VSPSTQSSQGKNADRSTLPKSVQEGNDSKCNAPSGKNGAAE-ANTDSPMKDLQGPAQNYD 215
Db    204 LLQIT----------------PQGTRDEISLDSAKKAACEFSETDVTNTEHHQPSNN-- 244

Qy    216 VAANVSEDNTSVDVGALPEVPQ-------ITWHIEVNGADQPPSTPKLSEVVLKRNEDEN 268
Db    245 -------DLNTTEKRAAERHPEKYQGSSVSNLHVEPCGTNTHASSLQHENSSLLLTKDRM 297

Qy    269 GKTEETLVAEQCNLTKDPNPMSGKERDQ------VAEQCNLTKDPKPVSGQK-----CEQ 317
Db    298 NVEK----AEFCNKSKQP----GLARSQHNRWAGSKETCN---DRRTPSTEKKVDLNADP 346

Qy    318 ICNE--------PCEE----------VVLKRSSKSKR----KTDKKLMKKQQHSKKRTAQ 355
Db    347 LCERKEWNKQKLPCSENPRDTEDVPWITLNSSIQKVNEWFSRSDELLGSDDSHDGESESN 406

Qy    356 ADVSDAKLCRRK-------PKKVRLLSEIINANQVEDSRSDEVHRENAADPCED------ 402
Db    407 AKVADVLDVLNEVDEYSGSSEKIDLLAS--DPHEALICKSDRVHSKSVESDIEDKIFGKT 464

Qy    403 --DRSTIPVPMEVSMDI---------------PVSNH-------TVG---EDGLKSSKNK 435
Db    465 YRKKASLPNLSHVTENLIIGAFVSEPQIIQERPLTNKLKRKRRPTSGLHPEDFIKKA-DL 523

Qy    436 TKRKYSDVVDDGSS-------LMNWLN-GKKKRT-------------------GSVHHTV 468
Db    524 AVQKTPEMINQGTNQTEQNGQVMNITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTK 583
```

```
Qy     469 AHP-AGNLSNKKVTPTASTQHDDENDTENGL-----DTNMHKTDVC-------QHVSEIS 515
           | | : ::||:      :  |: :   :| |       ::|  ::         : :|:
Db     584 AEPISSSISNEL---ELNIMHNSKAPKKNRLRRKSSTRHIHALELVVSRNLSPPNCTELQ 640

Qy     516 TQRCSSKGKTAGLSKGKTHSAASTKYGGESTRNGQNIHVLSAED------QCQMETENSV 569
             |||           :      ||      |: :|: ::  ::      :    | :
Db     641 IDSCSS----------SEEIKKKKYNQMPVRHSRNLQLMEGKEPATGAKKSNKPNEQTS 689

Qy     570 LSHSAKVSPAEHDIQI----------MSDLHE---QSLPKKKKKQKLEVTREKQTMID 614
             | :   |   ::::            |:| |    |||:::|::|||  :      |
Db     690 KRHDSDTFP---ELKLTNAPGSFTKCSNTSELKEFVNPSLPREEKEEKLETVKVSNNAED 746

Qy     615 DIPMDIVELLAKNQHERQLMTETDCSDINRIQSKTTADDDCVIVAAKDGSDY-------- 666
           | |::        || | ||        ::| :        ::   |:||
Db     747 --PKDLM-----LSGERVLQTE------RSVESSS--------ISLVPGTDYGTQESISL 785

Qy     667 --ASSVFDTNSQQKSLASQSTQKELQGHL--------ALTTQESPHPQNFQSTQEQQTHL 716
             |::      ::     ||    |     |          |:   :|   :    ::| :
Db     786 LEVSTLGKAKTEPNKCVSQCAAFENPKGLIHGCSKDNRNDTECFKYPLGHEVNHSRETSI 845

Qy     717 RMEEMVTIAASSPLFSHHDDQYIAE----------AP--------------TEHWG--RK 750
            |||           |  | ||:            ||              : | |  :|
Db     846 EMEE----------SELDAQYLQNTFKVSKRQSFAPFSNPGNAEEECATFSAHSGSLKK 894

Qy     751 DAKKLTWE-------QFKATTRNSPAAT---------CGAQFRPGIQAVDLTSTHVMGSS 794
           : |:|:|       | |  :   |  |          | : :|    ||      : | |
Db     895 QSPKVTFECEQKEENQGKNESNIKPVQTVNITAGFPVVGQKDKP----VDNAKCSIKGGS 950

Qy     795 SNYASRQ----------------------PVIAPLDRYAE-RAVNQVHARNFPSTIATM 830
             | |                        | : |:  : : :    :   ||
Db     951 RFCLSSQFRGNETGLITPNKHGLLQNPYRIPPLFPIKSFVKTKCKKNLLEENF------- 1003

Qy     831 EASKLCDRRNAGQVVLYPKESMPATHLLRMMDPSTLASFPNYGTSSRNQMESQLHNSQYA 890
           |   :   |  |       |::|:|                 | ||| :   :     :
Db    1004 EEHSMSPEREMG------NENIPST---------------VSTISRNNIRENVFKEASS 1041

Qy     891 HN-QYKGSTSTSYGSNLNGKIPLTFE----DLSRHQLHDLHRELRPHPRVGVLGSLLQKE 945
           |    ||::   ||::| :|  : |    :| |::   |:  ||       || :|| |
Db    1042 SNINEVGSSTNEVGSSIN-EIGSSDENIQAELGRNRGPKLNAMLR-------LG-VLQPE 1092

Qy     946 IANWSENCGTQSGYKLGVSTGITSHQMNRKEHFE 979
           :            ||  :      |   :|: :|
Db    1093 V-----------YKQSLPGSNCKHPEIKKQEYE 1114


RESULT   9
US-08-425-061-22
; Sequence 22, Application US/08425061
; Patent No. 5622829
;  GENERAL INFORMATION:
;    APPLICANT:  KING, Mary-Claire
;    APPLICANT:  FRIEDMAN, Lori
;    APPLICANT:  OSTERMEYER, Beth
;    APPLICANT:  ROWELL, Sarah
;    APPLICANT:  LYNCH, Eric
;    APPLICANT:  SZABO, Csilla
;    APPLICANT:  LEE, Ming
;    TITLE OF INVENTION:  GENETIC MARKERS FOR BREAST AND OVARIAN
;    TITLE OF INVENTION:  CANCER
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/425,061
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OSMAN, Richard A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  A-59563-3/DJB/RAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 494-8700
;      TELEFAX:  (415) 494-8771
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  22:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1202 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-425-061-22


  Query Match             3.2%;  Score 179;  DB 1;  Length 1202;
  Best Local Similarity   17.9%;  Pred. No. 3.1e-05;
  Matches  222;  Conservative  170;  Mismatches  381;  Indels  470;  Gaps   62;

Qy      48 VDEPAQHQCEHFSIRGYVALL--QKKDPKFCSLSRIFHDQKKCDEHKASSSPFS--VAKF 103
           : ||   :|:|   :  :  |  ||| |  | | :    ::   |    |: ||  | :
Db      31 IKEPVSTKCDHIFCKFCMLKLLNQKKGPSQCPLCKNDITKRSLQE----STRFSQLVEEL 86

Qy     104 RRWDCSKCLD---------------------------------------------------- 113
             :  |:  ||
Db      87 LKIICAFQLDTGLEYANSYNFAKKENNSPEHLKDEVSIIQSMGYRNRAKRLLQSEPENPS 146

Qy     114 ------KLKTSDNGTAPRTLPAKQNGT----------SDGCSITFVRSTFVPASVGSQK 156
                 :: |: ||  |||  ||              ||    |  ::|:   ||| |:
Db     147 LQETSLSVQLSNLGTV-RTLRTKQRIQPQKTSVYIELGSDSSEDTVNKATY--CSVGDQE 203

Qy     157 VSPSTQSSQGKNADRSTLPKSVQEGNDSKCNAPSGKNGAAE-ANTDSPMKDLQGPAQNYD 215
           :   |                   :|   : :  | |  | | : ||    :   |: |
Db     204 LLQIT----------------PQGTRDEISLDSAKKAACEFSETDVTNTEHHQPSNN-- 244

Qy     216 VAANVSEDNTSVDVGALPEVPQ-------ITWHIEVNGADQPPSTPKLSEVVLKRNEDEN 268
             |  : :  |    |:             |:|  | :   |: :      |   :|
Db     245 -------DLNTTEKRAAERHPEKYQGSSVSNLHVEPCGTNTHASSLQHENSSLLLTKDRM 297

Qy     269 GKTEETLVAEQCNLTKDPNPMSGKERDQ------VAEQCNLTKDPKPVSGQK-----CEQ 317
           :    || || :| |    |  | |         | ||   | :  | :|      :
Db     298 NVEK----AEFCNKSKQP----GLARSQHNRWAGSKETCN---DRRTPSTEKKVDLNADP 346

Qy     318 ICNE--------PCEE----------VVLKRSSKSKR----KTDKKLMKKQQHSKKRTAQ 355
           :|          || |          : :|  | :       ::|: |     |  :  :
Db     347 LCERKEWNKQKLPCSENPRDTEDVPWITLNSSIQKVNEWFSRSDELLGSDDSHDGESESN 406

Qy     356 ADVSD-------------------------AKLCRRKPKKVRLLSEIINANQVEDSRS 388
           | |:|                         | :|    |  |: |: : :| :||
Db     407 AKVADVLDVLNEVDEYSGSSEKIDLLASDPHEALIC----KSERVHSKSVESN-IEDKIF 461

Qy     389 DEVHRENAADPCEDDRSTIPVPM---------EVSMDIPVSNH-------TVG---EDGL 429
           : :|: |: |   : | :   :          ::  : |::|           | |   || :
Db     462 GKTYRKKASLP---NLSHVTENLIIGAFVTEPQIIQERPLTNKLKRKRRPTSGLHPEDFI 518

Qy     430 KSSKNKTKRKYSDVVDDGSS-------LMNWLN----GKKKRTGSVHHTVAHPAGNLSNK 478
           | : :   :|  :::: |::       :||  |     | |      :    :|  :|  :
Db     519 KKA-DLAVQKTPEMINQGTNQTEQNGQVMNITNSGHENKTKGDSIQNEKNPNPIESLEKE 577

Qy     479 -----KVTPTASTQHDDENDTENGLDTNMHKTDV----------CQHV---------- 511
                | | :|:  : |      |: |:| :                :|:
Db     578 SAFKTKAEPISSSISNME------LELNIHNSKAPKKNRLRRKSSTRHIHALELVVSRNL 631

Qy     512 -----SEISTQRCSSKGKTAGLSKGKTHSAASTKYGGESTRNGQNIHVLSAED------Q 560
                :|:    |||           :      ||     |: :|: ::  ::      :
Db     632 SPPNCTELQIDSCSS----------SEEIKKKKYNQMPVRHSRNLQLMEGKEPATGAKK 680
```

```
Qy     561 CQMETENSVLSHSAKVSPAEHDIQI----------MSDLHE---QSLPKKKKKQKLEV 605
Db     681 SNKPNEQTSKRHDSDTFP---ELKLTNAPGSFTKCSNTSELKEFVNPSLPREEKEEKLET 737

Qy     606 TREKQTMIDDIPMDIVELLAKNQHERQLMTETDCSDINRIQSKTTADDDCVIVAAKDGSD 665
Db     738 VKVSNNAED--PKDLM-----LSGERVLQTE------RSVESSS--------ISLVPGTD 776

Qy     666 Y----------ASSVFDTNSQQKSLASQSTQKELQGHL--------ALTTQESPHPQNFQ 707
Db     777 YGTQESISLLEVSTLGKAKTEPNKCVSQCAAFENPKGLIHGCSKDNRNDTEGFKYPLGHE 836

Qy     708 STQEQQTHLRMEEMVTIAASSPLFSHHDDQYIAE----------AP-------------- 743
Db     837 VNHSRETSIEMEE----------SELDAQYLQNTFKVSKRQSFAPFSNPGNAEEECATF 885

Qy     744 TEHWG--RKDAKKLTWE-------QFKATTRNSPAAT---------CGAQFRPGIQAVDL 785
Db     886 SAHSGSLKKQSPKVTFECEQKEENQGKNESNIKPVQTVNITAGFPVVGQKDKP----VDN 941

Qy     786 TSTHVMGSSSNYASRQ----------------------PVIAPLDRYAE-RAVNQVHAR 821
Db     942 AKCSIKGGSRFCLSSQFRGNETGLITPNKHGLLQNPYRIPPLFPIKSFVKTKCKKNLLEE 1001

Qy     822 NFPSTIATMEASKLCDRRNAGQVVLYPKESMPATHLLRMADPSTLASFPNYGTSSRNQME 881
Db    1002 NF-------EEHSMSPEREMG------NENIPST---------------VSTISRNNIR 1032

Qy     882 SQLHNSQYAHN-QYKGSTSTSYGSNLNGKIPLTFE----DLSRHQLHDLHRPLRPHPRVG 936
Db    1033 ENVFKEASSSNINEVGSSTNEVGSSIN-EIGSSDENIQAELGRNRGPKLNAMLR------ 1085

Qy     937 VLGSLLQKEIANWSENCGTQSGYKLGVSTGITSHQMNRKEHFE 979
Db    1086 -LG-VLQPEV----------YKQSLPGSNCKHPEIKKQEYE 1114
```

```
RESULT 10
US-08-825-886-22
; Sequence 22, Application US/08825886
; Patent No. 5821328
;  GENERAL INFORMATION:
;    APPLICANT:  KING, Mary-Claire
;    APPLICANT:  FRIEDMAN, Lori
;    APPLICANT:  OSTERMEYER, Beth
;    APPLICANT:  ROWELL, Sarah
;    APPLICANT:  LYNCH, Eric
;    APPLICANT:  SZABO, Csilla
;    APPLICANT:  LEE, Ming
;    TITLE OF INVENTION:  GENETIC MARKERS FOR BREAST AND OVARIAN
;    TITLE OF INVENTION:  CANCER
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/825,886
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/425,061
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OSMAN, Richard A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  A-59563-3/DJB/RAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 494-8700
;      TELEFAX:  (415) 494-8771
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  22:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1202 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-825-886-22
```

```
  Query Match             3.2%;  Score 179;  DB 2;  Length 1202;
  Best Local Similarity   17.9%;  Pred. No. 3.1e-05;
  Matches  222;  Conservative  170;  Mismatches  381;  Indels  470;  Gaps   62;

Qy      48 VDEPAQHQCEHFSIRGYVALL--QKKDPKFCSLSRIFHDQKKCDEHKASSSPFS--VAKF 103
Db      31 IKEPVSTKCDHIFCKFCMLKLLNQKKGPSQCPLCKNDITKRSLQE----STRFSQLVEEL 86

Qy     104 RRWDCSKCLD---------------------------------------------------- 113
Db      87 LKIICAFQLDTGLEYANSYNFAKKENNSPEHLKDEVSIIQSMGYRNRAKRLLQSEPENPS 146

Qy     114 ------KLKTSDNGTAPRTLPAKQNGT----------SDGCSITFVRSTFVPASVGSQK 156
Db     147 LQETSLSVQLSNLGTV-RTLRTKQRIQPQKTSVYIELGSDSSEDTVNKATY--CSVGDQE 203

Qy     157 VSPSTQSSQGKNADRSTLPKSVQEGNDSKCNAPSGKNGAAE-ANTDSPMKDLQGPAQNYD 215
Db     204 LLQIT-----------------PQGTRDEISLDSAKKAACEFSETDVTNTEHHQPSNN-- 244

Qy     216 VAANVSEDNTSVDVGALPEVPQ-------ITWHIEVNGADQPPSTPKLSEVVLKRNEDEN 268
Db     245 -------DLNTTEKRAAERHPEKYQGSSVSNLHVEPCGTNTHASSLQHENSSLLLTKDRM 297

Qy     269 GKTEETLVAEQCNLTKDPNPMSGKERDQ------VAEQCNLTKDPKPVSGQK-----CEQ 317
Db     298 NVEK----AEFCNKSKQP----GLARSQHNRWAGSKETCN---DRRTPSTEKKVDLNADP 346

Qy     318 ICNE--------PCEE----------VVLKRSSKSKR----KTDKKLMKKQQHSKKRTAQ 355
Db     347 LCERKEWNKQKLPCSENPRDTEDVPWITLNSSIQKVNEWFSRSDELLGSDDSHDGESESN 406

Qy     356 ADVSD--------------------------AKLCRRKPKKVRLLSEIINANQVEDSRS 388
Db     407 AKVADVLDVLNEVDEYSGSSEKIDLLASDPHEALIC----KSERVHSKSVESN-IEDKIF 461

Qy     389 DEVHRENAADPCEDDRSTIPVPM---------EVSMDIPVSNH-------TVG---EDGL 429
Db     462 GKTYRKKASLP---NLSHVTENLIIGAFVTEPQIIQERPLTNKLKRKRRPTSGLHPEDFI 518

Qy     430 KSSKNKTKRKYSDVVDDGSS-------LMNWLN----GKKKRTGSVHHTVAHPAGNLSNK 478
Db     519 KKA-DLAVQKTPEMINQGTNQTEQNGQVMNITNSGHENKTKGDSIQNEKNPNPIESLEKE 577

Qy     479 -----KVTPTASTQHDDENDTENGLDTNMHKTDV----------CQHV---------- 511
Db     578 SAFKTKAEPISSSISNME------LELNIHNSKAPKKNRLRRKSSTRHIHALELVVSRNL 631

Qy     512 -----SEISTQRCSSKGKTAGLSKGKTHSAASTKYGGESTRNGQNIHVLSAED------Q 560
Db     632 SPPNCTELQIDSCSS----------SEEIKKKKYNQMPVRHSRNLQLMEGKEPATGAKK 680

Qy     561 CQMETENSVLSHSAKVSPAEHDIQI----------MSDLHE---QSLPKKKKKQKLEV 605
Db     681 SNKPNEQTSKRHDSDTFP---ELKLTNAPGSFTKCSNTSELKEFVNPSLPREEKEEKLET 737

Qy     606 TREKQTMIDDIPMDIVELLAKNQHERQLMTETDCSDINRIQSKTTADDDCVIVAAKDGSD 665
```

```
          :      |  | |::        || | ||        ::| :        ::   |:|
Db    738 VKVSNNAED--PKDLM-----LSGERVLQTE------RSVESSS--------ISLVPGTD 776

Qy    666 Y----------ASSVFDTNSQQKSLASQSTQKELQGHL--------ALTTQESPHPQNFQ 707
          |           |::    ::     ||    |    |           |:   :|   :
Db    777 YGTQESISLLEVSTLGKAKTEPNKCVSQCAAFENPKGLIHGCSKDNRNDTEGFKYPLGHE 836

Qy    708 STQEQQTHLRMEEMVTIAASSPLFSHHDDQYIAE----------AP-------------- 743
              ::| : |||          |  | ||:           ||
Db    837 VNHSRETSIEMEE-----------SELDAQYLQNTFKVSKRQSFAPFSNPGNAEEECATF 885

Qy    744 TEHWG--RKDAKKLTWE-------QFKATTRNSPAAT---------CGAQFRPGIQAVDL 785
          : | |  :| : |:|:|       | |  :   |  |          | : :|    ||
Db    886 SAHSGSLKKQSPKVTFECEQKEENQGKNESNIKPVQTVNITACFPVVGQKDKP----VDN 941

Qy    786 TSTHVMGSSSNYASRQ----------------------PVIAPLDRYAE-RAVNQVHAR 821
              : | |    | |                        | : |:  : : :    :
Db    942 AKCSIKGGSRFCLSSQFRGNETGLITPNKHGLLQNPYRIPPLFPIKSFVKTKCKKNLLEE 1001

Qy    822 NFPSTIATMEASKLCDRRNAGQVVLYPKESMPATHLLRMMDPSTLASFPNYGTSSRNQME 881
          ||       |   :   |  |       |::|:|                    | ||| :
Db   1002 NF-------EEHSMSPEREMG------NENIPST---------------VSTISRNNIR 1032

Qy    882 SQLHNSQYAHN-QYKGSTSTSYGSNLNGKIPLTFE----DLSRHQLHDLHRPLRPHPRVG 936
            :    : |    ||::   ||::| :|  : |    :| |::   |:  ||
Db   1033 ENVFKEASSSNINEVGSSTNEVGSSIN-EIGSSDENIQAELGRNRGPKLNAMLR------ 1085

Qy    937 VLGSLLQKEIANWSENCGTQSGYKLGVSTGITSHQMNRKEHFE 979
            || :|| |:            ||  :      |   :|: :|
Db   1086 -LG-VLQPEV-----------YKQSLPGSNCKHPEIKKQEYE 1114


RESULT  11
US-08-425-061-23
; Sequence 23, Application US/08425061
; Patent No. 5622829
;  GENERAL INFORMATION:
;    APPLICANT:  KING, Mary-Claire
;    APPLICANT:  FRIEDMAN, Lori
;    APPLICANT:  OSTERMEYER, Beth
;    APPLICANT:  ROWELL, Sarah
;    APPLICANT:  LYNCH, Eric
;    APPLICANT:  SZABO, Csilla
;    APPLICANT:  LEE, Ming
;    TITLE OF INVENTION:  GENETIC MARKERS FOR BREAST AND OVARIAN
;    TITLE OF INVENTION:  CANCER
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/425,061
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OSMAN, Richard A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  A-59563-3/DJB/RAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 494-8700
;      TELEFAX:  (415) 494-8771
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1363 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-425-061-23


  Query Match           3.2%;  Score 179;  DB 1;  Length 1363;
  Best Local Similarity  17.9%;  Pred. No. 3.8e-05;
  Matches  222;  Conservative  170;  Mismatches  381;  Indels  470;  Gaps   62;

Qy     48 VDEPAQHQCEHFSIRGYVALL--QKKDPKFCSLSRIFHDQKKCDEHKASSSPFS--VAKF 103
          : ||   :|:|   :  :  |  ||| |  | | :    ::   |    |: ||  | :
Db     31 IKEPVSTKCDHIFCKFCMLKLLNQKKGPSQCPLCKNDITKRSLQE----STRFSQLVEEL 86

Qy    104 RRWDCSKCLD---------------------------------------------------- 113
          :  |:  ||
Db     87 LKIICAFQLDTGLEYANSYNFAKKENNSPEHLKDEVSIIQSMGYRNRAKRLLQSEPENPS 146

Qy    114 ------KLKTSDNGTAPRTLPAKQNGT----------SDGCSITFVRSTFVPASVGSQK 156
                :: |: ||  |||  ||                ||    |  ::|:   ||| |:
Db    147 LQETSLSVQLSNLGTV-RTLRTKQRIQPQKTSVYIELGSDSSEDTVNKATY--CSVGDQE 203

Qy    157 VSPSTQSSQGKNADRSTLPKSVQEGNDSKCNAPSGKNGAAE-ANTDSPMKDLQGPAQNYD 215
          :  |                  :|   : :  | |  | | : ||    :   |: |
Db    204 LLQIT----------------PQGTRDEISLDSAKKAACEFSETDVTNTEHHQPSNN-- 244

Qy    216 VAANVSEDNTSVDVGALPEVPQ-------ITWHIEVNGADQPPSTPKLSEVVLKRNEDEN 268
              |  : :  |    |:          |:|  | :   |: :     |   :|
Db    245 -------DLNTTEKRAAERHPEKYQGSSVSNLHVEPCGTNTHASSLQHENSSLLLTKDRM 297

Qy    269 GKTEETLVAEQCNLTKDPNPMSGKERDQ------VAEQCNLTKDPKPVSGQK-----CEQ 317
             :    || || :| |    |  | |         | ||   | :  | :|      :
Db    298 NVEK----AEFCNKSKQP----GLARSQHNRWAGSKETCN---DRRTPSTEKKVDLNADP 346

Qy    318 ICNE--------PCEE----------VVLKRSSKSKR----KTDKKLMKKQQHSKKRTAQ 355
          :|          || |          : |  | :        ::|: |     |  :  :
Db    347 LCERKEWNKQKLPCSENPRDTEDVPWITLNSSIQKVNEWFSRSDELLGSDDSHDGESESN 406

Qy    356 ADVSD-------------------------AKLCRRKPKKVRLLSEIINANQVEDSRS 388
          | |:|                         | :|    |  |: |: : :| :||
Db    407 AKVADVLDVLNEVDEYSGSSEKIDLLASDPHEALIC----KSERVHSKSVESN-IEDKIF 461

Qy    389 DEVHRENAADPCEDDRSTIPVPM---------EVSMDIPVSNH-------TVG---EDGL 429
           : :|: |: |   : | :   :          ::  : |::|         | |   || :
Db    462 GKTYRKKASLP---NLSHVTENLIIGAFVTEPQIIQERPLTNKLKRKRRPTSGLHPEDFI 518

Qy    430 KSSKNKTKRKYSDVVDDGSS-------LMNWLN----GKKKRTGSVHHTVAHPAGNLSNK 478
          | : :   :|  :::: |::         :||  |      | |      :   :|  :|  :
Db    519 KKA-DLAVQKTPEMINQGTNQTEQNGQVMNITNSGHENKTKGDSIQNEKNPNPIESLEKE 577

Qy    479 -----KVTPTASTQHDDENDTENGLDTNMHKTDV----------CQHV---------- 511
               |  :|:  : |       |: |:| :              :|:
Db    578 SAFKTKAEPISSSISNME------LELNIHNSKAPKKNRLRRKSSTRHIHALELVVSRNL 631

Qy    512 -----SEISTQRCSSKGKTAGLSKGKTHSAASTKYGGESTRNGQNIHVLSAED------Q 560
               :|:    |||           :      ||      |: :|: ::  ::      :
Db    632 SPPNCTELQIDSCSS----------SEEIKKKKYNQMPVRHSRNLQLMEGKEPATGAKK 680

Qy    561 CQMETENSVLSHSAKVSPAEHDIQI-----------MSDLHE---QSLPKKKKKQKLEV 605
              | :   | :   |   ::::                |:| |    |||:::|::|||
Db    681 SNKPNEQTSKRHDSDTFP---ELKLTNAPGSFTKCSNTSELKEFVNPSLPREEKEEKLET 737

Qy    606 TREKQTMIDDIPMDIVELLAKNQHERQLMTETDCSDINRIQSKTTADDDCVIVAAKDGSD 665
           :       |  | |::        || | ||        ::| :        ::   |:|
Db    738 VKVSNNAED--PKDLM-----LSGERVLQTE------RSVESSS--------ISLVPGTD 776

Qy    666 Y----------ASSVFDTNSQQKSLASQSTQKELQGHL--------ALTTQESPHPQNFQ 707
          |           |::    ::     ||    |    |           |:   :|   :
Db    777 YGTQESISLLEVSTLGKAKTEPNKCVSQCAAFENPKGLIHGCSKDNRNDTEGFKYPLGHE 836

Qy    708 STQEQQTHLRMEEMVTIAASSPLFSHHDDQYIAE----------AP-------------- 743
```

```
              ::| : |||           |  | ||:           ||
Db    837 VNHSRETSIEMEE----------SELDAQYLQNTFKVSKRQSFAPFSNPGNAEEECATF 885

Qy    744 TEHWG--RKDAKKLTWE-------QFKATTRNSPAAT---------CGAQFRPGIQAVDL 785
          : | |  :| : |:|:|       | |  :   |  |          | : :|    ||
Db    886 SAHSGSLKKQSPKVTFECEQKEENQGKNESNIKPVQTVNITAGFPVVGQKDKP----VDN 941

Qy    786 TSTHVMGSSSNYASRQ----------------------PVIAPLDRYAE-RAVNQVHAR 821
              : | |    | |                      | : |:  : : :    :
Db    942 AKCSIKGGSRFCLSSQFRGNETGLITPNKHGLLQNPYRIPPLFPIKSFVKTKCKKNLLEE 1001

Qy    822 NFPSTIATMEASKLCDRRNAGQVVLYPKESMPATHLLRMDPSTLASFPNYGTSSRNQME 881
          ||     |  :   |  |       |::|:|                  | ||| :
Db   1002 NF-------EEHSMSPEREMG------NENIPST----------------VSTISRNNIR 1032

Qy    882 SQLHNSQYAHN-QYKGSTSTSYGSNLNGKIPLTFE----DLSRHQLHDLHRPLRPHPRVG 936
            :    : |    ||::   ||::| :|  : |    :| |::   |:  ||
Db   1033 ENVFKEASSSNINEVGSSTNEVGSSIN-EIGSSDENIQAELGRNRGPKLNAMLR------ 1085

Qy    937 VLGSLLQKEIANWSENCGTQSGYKLGVSTGITSHQMNRIEHFE 979
            || :|| |:            ||  :      |   :|: :|
Db   1086 -LG-VLQPEV-----------YKQSLPGSNCKHPEIKIQEYE 1114


RESULT  12
US-08-825-886-23
; Sequence 23, Application US/08825886
; Patent No. 5821328
;  GENERAL INFORMATION:
;    APPLICANT:  KING, Mary-Claire
;    APPLICANT:  FRIEDMAN, Lori
;    APPLICANT:  OSTERMEYER, Beth
;    APPLICANT:  ROWELL, Sarah
;    APPLICANT:  LYNCH, Eric
;    APPLICANT:  SZABO, Csilla
;    APPLICANT:  LEE, Ming
;    TITLE OF INVENTION:  GENETIC MARKERS FOR BREAST AND OVARIAN
;    TITLE OF INVENTION:  CANCER
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/825,886
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/425,061
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OSMAN, Richard A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  A-59563-3/DJB/RAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 494-8700
;      TELEFAX:  (415) 494-8771
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1363 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein

US-08-825-886-23

  Query Match             3.2%;  Score 179;  DB 2;  Length 1363;
  Best Local Similarity   17.9%;  Pred. No. 3.8e-05;
  Matches  222;  Conservative  170;  Mismatches  381;  Indels  470;  Gaps   62;

Qy     48 VDEPAQHQCEHFSIRGYVALL--QKKDPKFCSLSRIFHDQKKCDEHKASSSPFS--VAKF 103
          : ||   :|:|   :  :  |  |||  |  | :    ::   |    |: ||  | :
Db     31 IKEPVSTKCDHIFCKFCMLKLLNQKKGPSQCPLCKNDITKRSLQE----STRFSQLVEEL 86

Qy    104 RRWDCSKCLD---------------------------------------------------- 113
           :  |:  ||
Db     87 LKIICAFQLDTGLEYANSYNFAKKENNSPEHLKDEVSIIQSMGYRNRAKRLLQSEPENPS 146

Qy    114 ------KLKTSDNGTAPRTLPAKQNGT----------SDGCSITFVRSTFVPASVGSQK 156
                ::  |: ||  |||  ||              ||    |  ::|:   ||| |:
Db    147 LQETSLSVQLSNLGTV-RTLRTKQRIQPQKTSVYIELGSDSSEDTVNKATY--CSVGDQE 203

Qy    157 VSPSTQSSQGKNADRSTLPKSVQEGNDSKCNAPSGKNGAAE-ANTDSPMKDLQGPAQNYD 215
          :  |                  :|   : :  | |  | | : ||    :   |: |
Db    204 LLQIT----------------PQGTRDEISLDSAKKAACEFSETDVTNTEHHQPSNN-- 244

Qy    216 VAANVSEDNTSVDVGALPEVPQ-------ITWHIEVNGADQPPSTPKLSEVVLKRNEDEN 268
                |  : :  |    |:          |:|  | :   |: :     |   :|
Db    245 -------DLNTTEKRAAERHPEKYQGSSVSNLHVEPCGTNTHASSLQHENSSLLLTKDRM 297

Qy    269 GKTEETLVAEQCNLTKDPNPMSGKERDQ------VAEQCNLTKDPKPVSGQK-----CEQ 317
             :    || || :| |    |  | |         | ||   | :  | :|     :
Db    298 NVEK----AEFCNKSKQP----GLARSQHNRWAGSKETCN---DRRTPSTEKKVDLNADP 346

Qy    318 ICNE--------PCEE----------VVLKRSSKSKR----KTDKKLMKKQQHSKKRTAQ 355
          :|          || |          : |  | :       ::|: |     |  :  :
Db    347 LCERKEWNKQKLPCSENPRDTEDVPWITLNSSIQKVNEWFSRSDELLGSDDSHDGESESN 406

Qy    356 ADVSD-------------------------AKLCRRKPKKVRLLSEIINANQVEDSRS 388
          | |:|                         | :|    |  |: |: : :| :||
Db    407 AKVADVLDVLNEVDEYSGSSEKIDLLASDPHEALIC----KSERVHSKSVESN-IEDKIF 461

Qy    389 DEVHRENAADPCEDDRSTIPVPM---------EVSMDIPVSNH-------TVG---EDGL 429
          : :|: |: |   : | :   :          ::  : |::|         | |   || :
Db    462 GKTYRKKASLP---NLSHVTENLIIGAFVTEPQIIQERPLTNKLKRKRRPTSGLHPEDFI 518

Qy    430 KSSKNKTKRKYSDVVDDGSS-------LMNWLN----GKKKRTGSVHHTVAHPAGNLSNK 478
          | : :   :|  :::: |::        :||  |      | |      :    :|  :|  :
Db    519 KKA-DLAVQKTPEMINQGTNQTEQNGQVMNITNSGHENKTKGDSIQNEKNPNPIESLEKE 577

Qy    479 -----KVTPTASTQHDDENDTENGLDTNMHKTDV----------CQHV---------- 511
               |  | :|:  : |       |: |:|  :             :|:
Db    578 SAFKTKAEPISSSISNME------LELNIHNSKAPKKNRLRRKSSTRHIHALELVVSRNL 631

Qy    512 -----SEISTQRCSSKGKTAGLSKGKTHSAASTKYGGESTRNGQNIHVLSAED------Q 560
               :|:   |||          :     ||    |: :|: ::  ::      :
Db    632 SPPNCTELQIDSCSS----------SEEIKKKKYNQMPVRHSRNLQLMEGKEPATGAKK 680

Qy    561 CQMETENSVLSHSAKVSPAEHDIQI----------MSDLHE---QSLPKKKKKQKLEV 605
             | :   | :   |   ::::            |:| |    |||:::|::|||
Db    681 SNKPNEQTSKRHDSDTFP---ELKLTNAPGSFTKCSNTSELKEFVNPSLPREEKEEKLET 737

Qy    606 TREKQTMIDDIPMDIVELLAKNQHERQLMTETDCSDINRIQSKTTADDDCVIVAAKDGSD 665
          :      |  | |::        || | ||        ::| :        ::   |:|
Db    738 VKVSNNAED--PKDLM-----LSGERVLQTE------RSVESSS--------ISLVPGTD 776

Qy    666 Y----------ASSVFDTNSQQKSLASQSTQKELQGHL--------ALTTQESPHPQNFQ 707
          |          |::    ::    ||   |   |           |:  :|   :
Db    777 YGTQESISLLEVSTLGKAKTEPNKCVSQCAAFENPKGLIHGCSKDNRNDTEGFKYPLGHE 836

Qy    708 STQEQQTHLRMEEMVTIAASSPLFSHHDDQYIAE----------AP-------------- 743
             ::| : |||           |  | ||:            ||
Db    837 VNHSRETSIEMEE----------SELDAQYLQNTFKVSKRQSFAPFSNPGNAEEECATF 885

Qy    744 TEHWG--RKDAKKLTWE-------QFKATTRNSPAAT---------CGAQFRPGIQAVDL 785
          : | |  :| : |:|:|       | |  :   |  |          | : :|    ||
```

```
Db    886 SAHSGSLKKQSPKVTFECEQKEENQGKNESNIKPVQTVNITAGFPVVGQKDKP----VDN 941

Qy    786 TSTHVMGSSSNYASRQ----------------------PVIAPLDRYAE-RAVNQVHAR 821
          : | |    | |                          | : |:  : : :    :
Db    942 AKCSIKGGSRFCLSSQFRGNETGLITPNKHGLLQNPYRIPPLFPIKSFVKTKCKKNLLEE 1001

Qy    822 NFPSTIATMEASKLCDRRNAGQVVLYPKESMPATHLLRMMDPSTLASFPNYGTSSRNQME 881
          ||       |   :   | |       |::|:|                  | ||| :
Db   1002 NF-------EEHSMSPEREMG------NENIPST---------------VSTISRNNIR 1032

Qy    882 SQLHNSQYAHN-QYKGSTSTSYGSNLNGKIPLTFE----DLSRHQLHDLHRPLRPHPRVG 936
          :    : |    ||::   ||::| :|  : |    :| |::   |:  ||
Db   1033 ENVFKEASSSNINEVGSSTNEVGSSIN-EIGSSDENIQAELGRNRGPKLNAMLR------ 1085

Qy    937 VLGSLLQKEIANWSENCGTQSGYKLGVSTGITSHQMNRKEHFE 979
            || :|| |:            ||  :      |   :|: :|
Db   1086 -LG-VLQPEV-----------YKQSLPGSNCKHPEIKKQEYE 1114
```

```
RESULT 13
US-08-425-061-24
; Sequence 24, Application US/08425061
; Patent No. 5622829
;  GENERAL INFORMATION:
;    APPLICANT:  KING, Mary-Claire
;    APPLICANT:  FRIEDMAN, Lori
;    APPLICANT:  OSTERMEYER, Beth
;    APPLICANT:  ROWELL, Sarah
;    APPLICANT:  LYNCH, Eric
;    APPLICANT:  SZABO, Csilla
;    APPLICANT:  LEE, Ming
;    TITLE OF INVENTION:  GENETIC MARKERS FOR BREAST AND OVARIAN
;    TITLE OF INVENTION:  CANCER
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/425,061
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OSMAN, Richard A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  A-59563-3/DJB/RAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 494-8700
;      TELEFAX:  (415) 494-8771
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1852 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-425-061-24


  Query Match             3.2%;  Score 179;  DB 1;  Length 1852;
  Best Local Similarity   17.9%;  Pred. No. 6.2e-05;
  Matches  222;  Conservative  170;  Mismatches  381;  Indels  470;  Gaps   62;

Qy     48 VDEPAQHQCEHFSIRGYVALL--QKKDPKFCSLSRIFHDQKKCDEHKASSSPFS--VAKF 103
          : ||   :|:|   :  :  |  ||| |  | | :    ::   |    |: ||  | :
Db     31 IKEPVSTKCDHIFCKFCMLKLLNQKKGPSQCPLCKNDITKRSLQE----STRFSQLVEEL 86

Qy    104 RRWDCSKCLD---------------------------------------------------- 113
          :  |:  ||
Db     87 LKIICAFQLDTGLEYANSYNFAKKENNSPEHLKDEVSIIQSMGYRNRAKRLLQSEPENPS 146

Qy    114 ------KLKTSDNGTAPRTLPAKQNGT-----------SDGCSITFVRSTFVPASVGSQK 156
               :: |: ||  |||  ||               ||    |  ::|:   ||| |:
Db    147 LQETSLSVQLSNLGTV-RTLRTKQRIQPQKTSVYIELGSDSSEDTVNKATY--CSVGDQE 203

Qy    157 VSPSTQSSQGKNADRSTLPKSVQEGNDSKCNAPSGKNGAAE-ANTDSPMKDLQGPAQNYD 215
          :   |                  :|   : :  | |  | | : ||    :   |: |
Db    204 LLQIT----------------PQGTRDEISLDSAKKAACEFSETDVTNTEHHQPSNN-- 244

Qy    216 VAANVSEDNTSVDVGALPEVPQ-------ITWHIEVNGADQPPSTPKLSEVVLKRNEDEN 268
                |  : :  |    |:          |:|  | :   |: :     |   :|
Db    245 -------DLNTTEKRAAERHPEKYQGSSVSNLHVEPCGTNTHASSLQHENSSLLLTKDRM 297

Qy    269 GKTEETLVAEQCNLTKDPNPMSGKERDQ------VAEQCNLTKDPKPVSGQK-----CEQ 317
          :    || || :| |    |  | |         | ||   | :  | :|      :
Db    298 NVEK----AEFCNKSKQP----GLARSQHNRWAGSKETCN---DRRTPSTEKKVDLNADP 346

Qy    318 ICNE--------PCEE----------VVLKRSSKSKR----KTDKKLMKKQQHSKKRTAQ 355
          :|          || |          : |  | :       ::|: |     |  :  :
Db    347 LCERKEWNKQKLPCSENPRDTEDVPWITLNSSIQKVNEWFSRSDELLGSDDSHDGESESN 406

Qy    356 ADVSD-------------------------AKLCRRKPKKVRLLSEIINANQVEDSRS 388
          | |:|                         | :|    |  |: |: : :| :||
Db    407 AKVADVLDVLNEVDEYSGSSEKIDLLASDPHEALIC----KSERVHSKSVESN-IEDKIF 461

Qy    389 DEVHRENAADPCEDDRSTIPVPM---------EVSMDIPVSNH-------TVG---EDGL 429
          : :|: |: |   : | :   :          ::  : |::|         | |   || :
Db    462 GKTYRKKASLP---NLSHVTENLIIGAFVTEPQIIQERPLTNKLKRKRRPTSGLHPEDFI 518

Qy    430 KSSKNKTKRKYSDVVDDGSS-------LMNWLN----GKKKRTGSVHHTVAHPAGNLSNK 478
          | : :   :|  :::: |::       :||  |     | |      :   :|  :|  :
Db    519 KKA-DLAVQKTPEMINQGTNQTEQNGQVMNITNSGHENKTKGDSIQNEKNPNPIESLEKE 577

Qy    479 -----KVTPTASTQHDDENDTENGLDTNMHKTDV----------CQHV---------- 511
               |  | :|:  : |       |: |:| :           :|:
Db    578 SAFKTKAEPISSSISNME------LELNIHNSKAPKKNRLRRKSSTRHIHALELVVSRNL 631

Qy    512 -----SEISTQRCSSKGKTAGLSKGKTHSAASTKYGGESTRNGQNIHVLSAED------Q 560
               :|:    |||          :      ||     |: :|: :: ::       :
Db    632 SPPNCTELQIDSCSS----------SEEIKKKKYNQMPVRHSRNLQLMEGKEPATGAKK 680

Qy    561 CQMETENSVLSHSAKVSPAEHDIQI----------MSDLHE---QSLPKKKKKQKLEV 605
               | :   | :   |   ::::             |:| |    |||:::|::|||
Db    681 SNKPNEQTSKRHDSDTFP---ELKLTNAPGSFTKCSNTSELKEFVNPSLPREEKEEKLET 737

Qy    606 TREKQTMIDDIPMDIVELLAKNQHERQLMTETDCSDINRIQSKTTADDDCVIVAAKDGSD 665
          :      |  | |::        || | ||        ::| :         ::   |:|
Db    738 VKVSNNAED--PKDLM-----LSGERVLQTE------RSVESSS--------ISLVPGTD 776

Qy    666 Y----------ASSVFDTNSQQKSLASQSTQKELQGHL--------ALTTQESPHPQNFQ 707
          |          |::    ::    ||    |    |            |:   :|   :
Db    777 YGTQESISLLEVSTLGKAKTEPNKCVSQCAAFENPKGLIHGCSKDNRNDTEGFKYPLGHE 836

Qy    708 STQEQQTHLRMEEMVTIAASSPLFSHHDDQYIAE----------AP---------- 743
             ::| : |||           |  | ||:           ||
Db    837 VNHSRETSIEMEE----------SELDAQYLQNTFKVSKRQSFAPFSNPGNAEEECATF 885

Qy    744 TEHWG--RKDAKKLTWE-------QFKATTRNSPAAT---------CGAQFRPGIQAVDL 785
          : | |  :| : |:|:|       | |  :   |  |           | :|    ||
Db    886 SAHSGSLKKQSPKVTFECEQKEENQGKNESNIKPVQTVNITAGFPVVGQKDKP----VDN 941

Qy    786 TSTHVMGSSSNYASRQ----------------------PVIAPLDRYAE-RAVNQVHAR 821
          : | |    | |                          | : |:  : : :    :
Db    942 AKCSIKGGSRFCLSSQFRGNETGLITPNKHGLLQNPYRIPPLFPIKSFVKTKCKKNLLEE 1001

Qy    822 NFPSTIATMEASKLCDRRNAGQVVLYPKESMPATHLLRMMDPSTLASFPNYGTSSRNQME 881
          ||       |   :   | |       |::|:|                  | ||| :
```

```
Db   1002 NF-------EEHSMSPEREMG------NENIPST----------------VSTISRNNIR 1032

Qy    882 SQLHNSQYAHN-QYKGSTSTSYGSNLNGKIPLTFE----DLSRHQLHDLHRPLRPHPRVG 936
Db   1033 ENVFKEASSSNINEVGSSTNEVGSSIN-EIGSSDENIQAELGRNRGPKLNAMLR------ 1085

Qy    937 VLGSLLQKEIANWSENCGTQSGYKLGVSTGITSHQMNRKEHFE 979
Db   1086 -LG-VLQPEV------------YKQSLPGSNCKHPEIKKQEYE 1114


RESULT 14
US-08-825-886-24
; Sequence 24, Application US/08825886
; Patent No. 5821328
;  GENERAL INFORMATION:
;    APPLICANT:  KING, Mary-Claire
;    APPLICANT:  FRIEDMAN, Lori
;    APPLICANT:  OSTERMEYER, Beth
;    APPLICANT:  ROWELL, Sarah
;    APPLICANT:  LYNCH, Eric
;    APPLICANT:  SZABO, Csilla
;    APPLICANT:  LEE, Ming
;    TITLE OF INVENTION:  GENETIC MARKERS FOR BREAST AND OVARIAN
;    TITLE OF INVENTION:  CANCER
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/825,886
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/425,061
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OSMAN, Richard A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  A-59563-3/DJB/RAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 494-8700
;      TELEFAX:  (415) 494-8771
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1852 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-825-886-24


  Query Match             3.2%;  Score 179;  DB 2;  Length 1852;
  Best Local Similarity   17.9%;  Pred. No. 6.2e-05;
  Matches  222;  Conservative  170;  Mismatches  381;  Indels  470;  Gaps   62;

Qy      48 VDEPAQHQCEHFSIRGYVALL--QKKDPKFCSLSRIFHDQKKCDEHKASSSPFS--VAKF 103
Db      31 IKEPVSTKCDHIFCKFCMLKLLNQKKGPSQCPLCKNDITKRSLQE----STRFSQLVEEL 86

Qy     104 RRWDCSKCLD---------------------------------------------------- 113

Db      87 LKIICAFQLDTGLEYANSYNFAKKENNSPEHLKDEVSIIQSMGYRNRAKRLLQSEPENPS 146

Qy     114 ------KLKTSDNGTAPRTLPAKQNGT-----------SDGCSITFVRSTFVPASVGSQK 156
Db     147 LQETSLSVQLSNLGTV-RTLRTKQRIQPQKTSVYIELGSDSSEDTVNKATY--CSVGDQE 203

Qy     157 VSPSTQSSQGKNADRSTLPKSVQEGNDSKCNAPSGKNGAAE-ANTDSPMKDLQGPAQNYD 215
Db     204 LLQIT----------------PQGTRDEISLDSAKKAACEFSETDVTNTEHHQPSNN-- 244

Qy     216 VAANVSEDNTSVDVGALPEVPQ-------ITWHIEVNGADQPPSTPKLSEVVLKRNEDEN 268
Db     245 -------DLNTTEKRAAERHPEKYQGSSVSNLHVEPCGTNTHASSLQHENSSLLLTKDRM 297

Qy     269 GKTEETLVAEQCNLTKDPNPMSGKERDQ------VAEQCNLTKDPKPVSGQK-----CEQ 317
Db     298 NVEK----AEFCNKSKQP----GLARSQHNRWAGSKETCN---DRRTPSTEKKVDLNADP 346

Qy     318 ICNE--------PCEE----------VVLKRSSKSKR----KTDKKLMKKQQHSKKRTAQ 355
Db     347 LCERKEWNKQKLPCSENPRDTEDVPWITLNSSIQKVNEWFSRSDELLGSDDSHDGESESN 406

Qy     356 ADVSD--------------------------AKLCRRKPKKVRLLSEIINANQVEDSRS 388
Db     407 AKVADVLDVLNEVDEYSGSSEKIDLLASDPHEALIC----KSERVHSKSVESN-IEDKIF 461

Qy     389 DEVHRENAADPCEDDRSTIPVPM---------EVSMDIPVSNH-------TVG---EDGL 429
Db     462 GKTYRKKASLP---NLSHVTENLIIGAFVTEPQIIQERPLTNKLKRKRRPTSGLHPEDFI 518

Qy     430 KSSKNKTKRKYSDVVDDGSS-------LMNWLN----GKKKRTGSVHHTVAHPAGNLSNK 478
Db     519 KKA-DLAVQKTPEMINQGTNQTEQNGQVMNITNSGHENKTKGDSIQNEKNPNPIESLEKE 577

Qy     479 -----KVTPTASTQHDDENDTENGLDTNMHKTDV----------CQHV---------- 511
Db     578 SAFKTKAEPISSSISNME------LELNIHNSKAPKKNRLRRKSSTRHIHALELVVSRNL 631

Qy     512 -----SEISTQRCSSKGKTAGLSKGKTHSAASTKYGGESTRNGQNIHVLSAED------Q 560
Db     632 SPPNCTELQIDSCSS----------SEEIKKKKYNQMPVRHSRNLQLMEGKEPATGAKK 680

Qy     561 CQMETENSVLSHSAKVSPAEHDIQI----------MSDLHE---QSLPKKKKKQKLEV 605
Db     681 SNKPNEQTSKRHDSDTFP---ELKLTNAPGSFTKCSNTSELKEFVNPSLPREEKEEKLET 737

Qy     606 TREKQTMIDDIPMDIVELLAKNQHERQLMTETDCSDINRIQSKTTADDDCVIVAAKDGSD 665
Db     738 VKVSNNAED--PKDLM-----LSGERVLQTE------RSVESSS--------ISLVPGTD 776

Qy     666 Y----------ASSVFDTNSQQKSLASQSTQKELQGHL--------ALTTQESPHPQNFQ 707
Db     777 YGTQESISLLEVSTLGKAKTEPNKCVSQCAAFENPKGLIHGCSKDNRNDTEGFKYPLGHE 836

Qy     708 STQEQQTHLRMEEMVTIAASSPLFSHHDDQYIAE----------AP---------- 743
Db     837 VNHSRETSIEMEE----------SELDAQYLQNTFKVSKRQSFAPFSNPGNAEEECATF 885

Qy     744 TEHWG--RKDAKKLTWE-------QFKATTRNSPAAT---------CGAQFRPGIQAVDL 785
Db     886 SAHSGSLKKQSPKVTFECEQKEENQGKNESNIKPVQTVNITAGFPVVGQKDKP----VDN 941

Qy     786 TSTHVMGSSSNYASRQ----------------------PVIAPLDRYAE-RAVNQVHAR 821
Db     942 AKCSIKGGSRFCLSSQFRGNETGLITPNKHGLLQNPYRIPPLFPIKSFVKTKCKKNLLEE 1001

Qy     822 NFPSTIATMEASKLCDRRNAGQVVLYPKESMPATHLLRMMDPSTLASFPNYGTSSRNQME 881
Db    1002 NF-------EEHSMSPEREMG------NENIPST----------------VSTISRNNIR 1032

Qy     882 SQLHNSQYAHN-QYKGSTSTSYGSNLNGKIPLTFE----DLSRHQLHDLHRPLRPHPRVG 936
Db    1033 ENVFKEASSSNINEVGSSTNEVGSSIN-EIGSSDENIQAELGRNRGPKLNAMLR------ 1085
```

```
Qy    937 VLGSLLQKEIANWSENCGTQSGYKLGVSTGITSHQMNRKEHFE 979
          || :|| |:            ||  :      |   :|: :|
Db   1086 -LG-VLQPEV-----------YKQSLPGSNCKHPEIKKQEYE 1114


RESULT 15
US-08-480-784-2
; Sequence 2, Application US/08480784
; Patent No. 5693473
;  GENERAL INFORMATION:
;    APPLICANT:  Skolnick, Mark H.
;    APPLICANT:  Goldgar, David E.
;    APPLICANT:  Miki, Yoshio
;    APPLICANT:  Swenson, Jeff
;    APPLICANT:  Kamb, Alexander
;    APPLICANT:  Harshman, Keith D.
;    APPLICANT:  Shattuck-Eidens, Donna M.
;    APPLICANT:  Tavtigian, Sean V.
;    APPLICANT:  Wiseman, Roger W.
;    APPLICANT:  Futreal, P. Andrew
;    TITLE OF INVENTION:  17q-Linked Breast and Ovarian Cancer
;    TITLE OF INVENTION:  Susceptibility Gene
;    NUMBER OF SEQUENCES:  85
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Venable, Baetjer, Howard & Civiletti, LLP
;      STREET:  1201 New York Avenue, N.W., Suite 1000
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,784
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/409,305
;      FILING DATE:  24-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/348,824
;      FILING DATE:  29-NOV-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/308,104
;      FILING DATE:  16-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/300,266
;      FILING DATE:  02-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/289,221
;      FILING DATE:  12-AUG-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ihnen, Jeffrey L.
;      REGISTRATION NUMBER:  28,957
;      REFERENCE/DOCKET NUMBER:  24884-109347
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-962-4810
;      TELEFAX:  202-962-8300
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;     LENGTH: 1863 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-480-784-2


  Query Match             3.2%;  Score 179;  DB 1;  Length 1863;
  Best Local Similarity  17.9%;  Pred. No. 6.2e-05;

  Matches  222;  Conservative  170;  Mismatches  381;  Indels  470;  Gaps   62;

Qy     48 VDEPAQHQCEHFSIRGYVALL--QKKDPKFCSLSRIFHDQKKCDEHKASSSPFS--VAKF 103
          : ||   :|:|   :  :  |  ||| |  | | :   ::   |    |: ||  | :
Db     31 IKEPVSTKCDHIFCKFCMLKLLNQKKGPSQCPLCKNDITKRSLQE---STRFSQLVEEL 86

Qy    104 RRWDCSKCLD---------------------------------------------------- 113
           :  |:  ||
Db     87 LKIICAFQLDTGLEYANSYNFAKKENNSPEHLKDEVSIIQSMGYRNRAKRLLQSEPENPS 146

Qy    114 ------KLKTSDNGTAPRTLPAKQNGT-----------SDGCSITFVRSTFVPASVGSQK 156
                :: |: ||  |||  ||              ||    |  ::|:   ||| |:
Db    147 LQETSLSVQLSNLGTV-RTLRTKQRIQPQKTSVYIELGSDSSEDTVNKATY--CSVGDQE 203

Qy    157 VSPSTQSSQGKNADRSTLPKSVQEGNDSKCNAPSGKNGAAE-ANTDSPMKDLQGPAQNYD 215
          :   |                  :|   : :  | |  | | : ||    :   |: |
Db    204 LLQIT----------------PQGTRDEISLDSAKKAACEFSETDVTNTEHHQPSNN-- 244

Qy    216 VAANVSEDNTSVDVGALPEVPQ-------ITWHIEVNGADQPPSTPKLSEVVLKRNEDEN 268
                |  : :  |    |:          |:|  | :   |: :     |  :|
Db    245 -------DLNTTEKRAAERHPEKYQGSSVSNLHVEPCGTNTHASSLQHENSSLLLTKDRM 297

Qy    269 GKTEETLVAEQCNLTKDPNPMSGKERDQ------VAEQCNLTKDPKPVSGQK-----CEQ 317
          :    || || :| |    |  | |         | ||   | :  | :|     :
Db    298 NVEK----AEFCNKSKQP----GLARSQHNRWAGSKETCN---DRRTPSTEKKVDLNADP 346

Qy    318 ICNE--------PCEE----------VVLKRSSKSKR----KTDKKLMKKQQHSKKRTAQ 355
          :|          || |          : |  | :       ::|: |     |  :  :
Db    347 LCERKEWNKQKLPCSENPRDTEDVPWITLNSSIQKVNEWFSRSDELLGSDDSHDGESESN 406

Qy    356 ADVSD--------------------------AKLCRRKPKKVRLLSEIINANQVEDSRS 388
          | |:|                          | :|    |  |: |: : :| :||
Db    407 AKVADVLDVLNEVDEYSGSSEKIDLLASDPHEALIC----KSERVHSKSVESN-IEDKIF 461

Qy    389 DEVHRENAADPCEDDRSTIPVPM---------EVSMDIPVSNH-------TVG---EDGL 429
          : :|: |: |   : | :   :          ::  : |::|        | |   || :
Db    462 GKTYRKKASLP---NLSHVTENLIIGAFVTEPQIIQERPLTNKLKRKRRPTSGLHPEDFI 518

Qy    430 KSSKNKTKRKYSDVVDDGSS-------LMNWLN----GKKKRTGSVHHTVAHPAGNLSNK 478
          | : :   :|  :::: |::       :||  |     | |      :   :|  :|  :
Db    519 KKA-DLAVQKTPEMINQGTNQTEQNGQVMNITNSGHENKTKGDSIQNEKNPNPIESLEKE 577

Qy    479 -----KVTPTASTQHDDENDTENGLDTNMHKTDV----------CQHV---------- 511
               |  | :|:  : |      |: |:| :              :|:
Db    578 SAFKTKAEPISSSISNME------LELNIHNSKAPKKNRLRRKSSTRHIHALELVVSRNL 631

Qy    512 -----SEISTQRCSSKGKTAGLSKGKTHSAASTKYGGESTRNGQNIHVLSAED------Q 560
               :|:    |||           :      ||      |: :|: ::  ::      :
Db    632 SPPNCTELQIDSCSS----------SEEIKKKKYNQMPVRHSRNLQLMEGKEPATGAKK 680

Qy    561 CQMETENSVLSHSAKVSPAEHDIQI----------MSDLHE---QSLPKKKKKQKLEV 605
               | :   | :   |   ::::           |:| |    |||:::|::|||
Db    681 SNKPNEQTSKRHDSDTFP---ELKLTNAPGSFTKCSNTSELKEFVNPSLPREEKEEKLET 737

Qy    606 TREKQTMIDDIPMDIVELLAKNQHERQLMTETDCSDINRIQSKTTADDDCVIVAAKDGSD 665
           :      |  | |::        || | ||        ::| :        ::   |:|
Db    738 VKVSNNAED--PKDLM-----LSGERVLQTE------RSVESSS--------ISLVPGTD 776

Qy    666 Y----------ASSVFDTNSQQKSLASQSTQKELQGHL--------ALTTQESPHPQNFQ 707
          |          |::    ::     ||   |   |           |:   :|   :
Db    777 YGTQESISLLEVSTLGKAKTEPNKCVSQCAAFENPKGLIHGCSKDNRNDTEGFKYPLGHE 836

Qy    708 STQEQQTHLRMEEMVTIAASSPLFSHHDDQYIAE----------AP-------------- 743
             ::| : |||           |  | ||:             ||
Db    837 VNHSRETSIEMEE-----------SELDAQYLQNTFKVSKRQSFAPFSNPGNAEEECATF 885

Qy    744 TEHWG--RKDAKKLTWE-------QFKATTRNSPAAT---------CGAQFRPGIQAVDL 785
          : | |  :| : |:|:|        | |  :   |  |          | : :|   ||
Db    886 SAHSGSLKKQSPKVTFECEQKEENQGKNESNIKPVQTVNITAGFPVVGQKDKP----VDN 941

Qy    786 TSTHVMGSSSNYASRQ----------------------PVIAPLDRYAE-RAVNQVHAR 821
          : | |   | |                          | : |:  : : :    :
Db    942 AKCSIKGGSRFCLSSQFRGNETGLITPNKHGLLQNPYRIPPLFPIKSFVKTKCKKNLLEE 1001
```

```
Qy     822 NFPSTIATMEASKLCDRRNAGQVVLYPKESMPATHLLRMMDPSTLASFPNYGTSSRNQME 881
           ||       |   :   |  |       |::|:|                 | ||| :
Db    1002 NF-------EEHSMSPEREMG------NENIPST---------------VSTISRNNIR 1032

Qy     882 SQLHNSQYAHN-QYKGSTSTSYGSNLNGKIPLTFE----DLSRHQLHDLHRPLRPHPRVG 936
             :     : |    ||::   ||::| :|  : |    :| |::   |:  ||
Db    1033 ENVFKEASSSNINEVGSSTNEVGSSIN-EIGSSDENIQAELGRNRGPKLNAMLR------ 1085

Qy     937 VLGSLLQKEIANWSENCGTQSGYKLGVSTGITSHQMNRKEHFE 979
           || :|| |:           ||  :      |   :|: :|
Db    1086 -LG-VLQPEV----------YKQSLPGSNCKHPEIKKQEYE 1114
```

Search completed: September 16, 2002, 22:24:02
Job time: 7446 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: September 16, 2002, 20:36:51 ; Search time 50.66 Seconds
(without alignments)
2004.864 Million cell updates/sec

Title: US-09-828-068-2
Perfect score: 5526
Sequence: 1 MEIVAVDQEGARVVGTNCML.....NKNPADFTTISNDNEYMDYR 1057

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

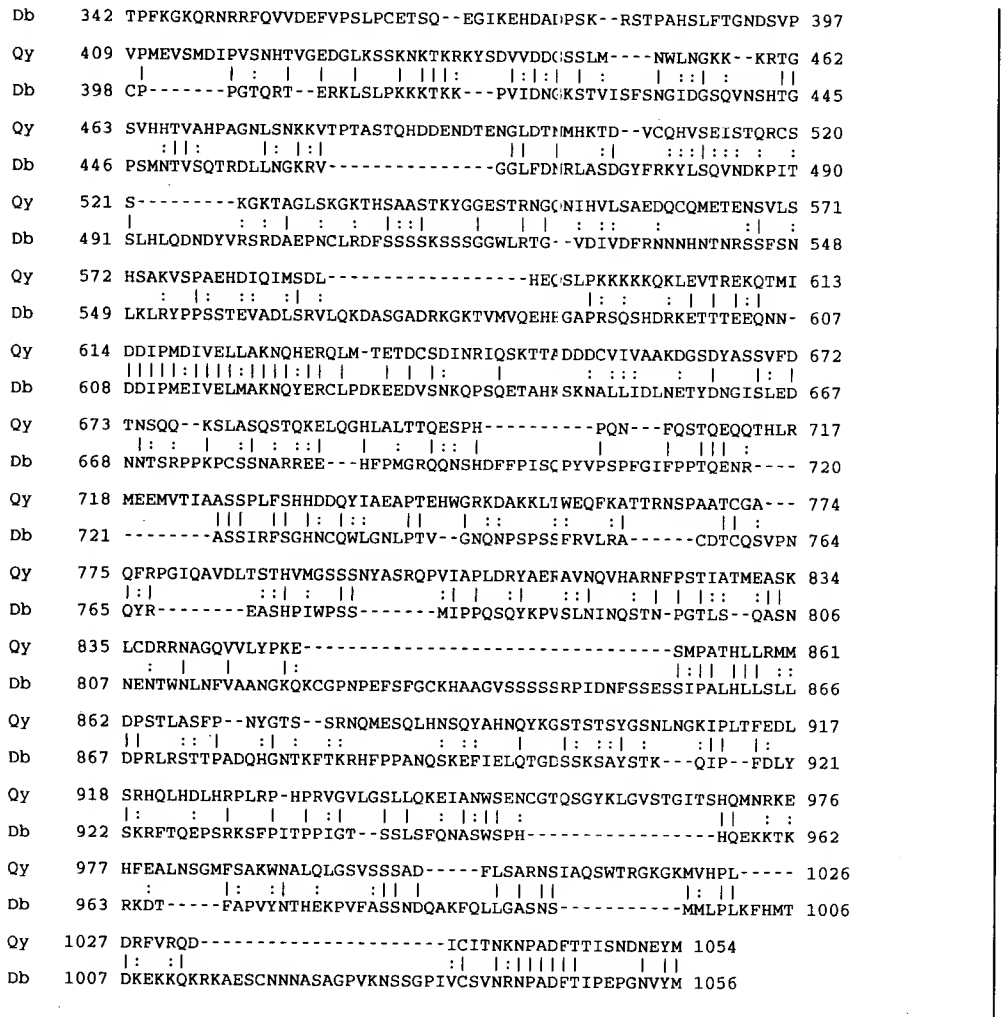
Result No.	Score	% Query Match	Length	DB	ID	Description
1	393	7.1	1096	2	T48512	hypothetical prote
2	225	4.1	1507	2	B47328	natural killer cel
3	221	4.0	1983	2	T00385	KIAA0624 protein -
4	217	3.9	1165	2	S62982	vacuolar protein V
5	215.5	3.9	3507	2	T34513	hypothetical prote
6	215	3.9	1690	2	T13030	microtubule bindin
7	213.5	3.9	1560	2	T42727	proliferation pote
8	209.5	3.8	852	2	T06310	hypothetical prote
9	208.5	3.8	1295	2	T24587	hypothetical prote
10	208	3.8	1630	2	A53577	ascites sialoglyco
11	207.5	3.8	1520	2	T44231	hypothetical prote
12	204.5	3.7	5327	2	T13564	microtubule-associ
13	204	3.7	3488	2	T34418	hypothetical prote
14	202.5	3.7	2271	2	F90073	hypothetical prote
15	201.5	3.6	665	2	B71609	hypothetical prote
16	200.5	3.6	1849	2	C41859	IgA-specific metal
17	199.5	3.6	1302	1	JC6009	surface-located me
18	199	3.6	1230	2	T22458	hypothetical prote
19	198.5	3.6	1043	2	D84900	hypothetical prote
20	196	3.5	1513	2	T44045	hypothetical prote
21	195	3.5	1641	2	I38614	helicase II - huma
22	194.5	3.5	2738	2	E88320	protein F07A11.6 [
23	194.5	3.5	3924	2	S37431	ankyrin 2, neurona
24	192	3.5	6642	2	T29757	protein UNC-89 - C
25	191.5	3.5	990	2	I51618	nucleolar phosphop
26	191	3.5	1233	2	S56271	hypothetical prote
27	188	3.4	1576	2	S65774	homeotic protein H
28	188	3.4	1829	2	T24583	hypothetical prote
29	187.5	3.4	904	2	T22457	hypothetical prote
30	187.5	3.4	1234	2	T31623	hypothetical prote
31	187.5	3.4	2722	2	T20532	hypothetical prote
32	186.5	3.4	1871	2	D96796	probable heat shoc
33	186	3.4	1269	2	F84730	probable myosin he
34	185.5	3.4	1080	2	T43164	IactA protein - Li
35	185	3.3	1061	2	C88690	protein F41H10.4 [
36	184.5	3.3	962	2	T22459	hypothetical prote
37	184.5	3.3	1091	2	T34107	hypothetical prote
38	184	3.3	2364	2	A56577	microtubule-associ
39	183.5	3.3	1459	2	T32271	hypothetical prote
40	183	3.3	875	2	S70115	ZIP1 protein - yea
41	183	3.3	1224	2	F96795	hypothetical prote
42	183	3.3	1927	2	A59236	embryonic muscle m
43	183	3.3	2481	2	D90011	FmtB protein [impo
44	182.5	3.3	990	2	H88733	protein F32E10.3 [
45	182.5	3.3	1032	2	T30270	major tegumental a

ALIGNMENTS

RESULT 1
T48512
hypothetical protein F15N18.120 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 05-May-2000
C;Accession: T48512
R;Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancro
submitted to the Protein Sequence Database, April 2000
A;Reference number: Z24490
A;Accession: T48512
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1096 <BEV>
A;Cross-references: EMBL:AL163815
A;Experimental source: cultivar Columbia; BAC clone F15N18
C;Genetics:
A;Map position: 5
A;Introns: 32/2; 91/1; 131/1; 149/1; 569/3
A;Note: F15N18.120
C;Superfamily: Arabidopsis thaliana hypothetical protein F15N18.120

Query Match 7.1%; Score 393; DB 2; Length 1096;
Best Local Similarity 21.0%; Pred. No. 1.7e-13;
Matches 250; Conservative 189; Mismatches 403; Indels 348; Gaps 58;

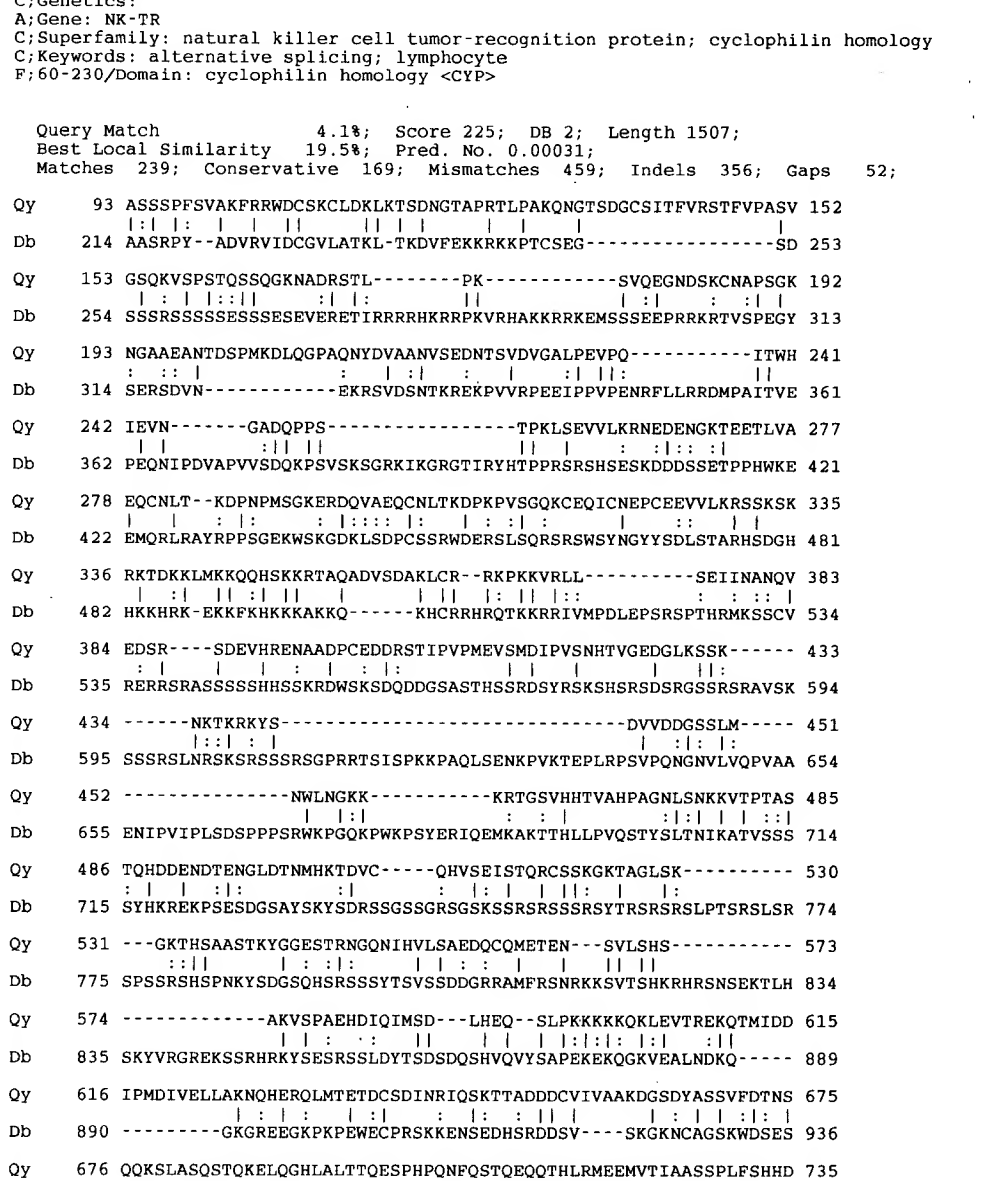
```
Qy      55 QCEHFSIRGYVALLQKKDPKFC------SLSRIFHDQKKCDEHKASSSPFSVAKFRRWDC 108
           :|:|||:||:||  :::| : |      |:| :       |:   :    || ||| | |
Db      25 KCDHFSMRGFVAETRERDLRKCWPFSEESVSLV-------DQQSYTLPTLSVPKFRWWHC 77

Qy     109 SKCLDKLKTSDNGTAPRTLPAKQNGTSDG-CSITFVRSTFVPASVGSQKVSPSTQSSQGK 167
             |:  :    :   |:       |  : |  |:   :| |   ::   :    |
Db      78 MSCIKDI----DAHGPKDCGLHSNSKAIGNSSVIESKSKFNSLTIIDHEKEKKTDI---- 129

Qy     168 NADRSTLPK-SVQEGNDSKC---------NAPSGKNGAAEANTDSPMKDLQGPAQNYDVA 217
           || :   |  |   || :            | |      :|  |  : |  | |   |
Db     130 -ADNAIEEKVGVNCENDDQTATTFLKKARGRPMG-----ASNVRSKSRKLVSPEQ---VG 180

Qy     218 ANVSED---NTSVDVGALPEVPQITWHIEVNGA--------DQPPSTPKLSEVVLKRNED 266
           | |::     |:|: :  |   :   :   |:        | ||   |  : :    |
Db     181 NNRSKEKLNKPSMDISSWKEKQNVDQAVTTFGSSEIAGVVEDTPPKATKNHKGIRGLMEC 240

Qy     267 ENGKTEETLVAEQCNLTKDPNPMSGKER---DQVAEQCNLTKDPKPVSGQKCEQICNEPC 323
           :|| :|   :|           ||| :|    :|     |  : |   |    :
Db     241 DNGSSESINLA----------MSGLQRRKSRKVRLLSELLGNTKTSGGSNIRK------ 283

Qy     324 EEVVLKRSSKSKRKTDKKLMKKQQH-SKKRTAQADVSD--AKLC---------------- 364
           ||  ||: |   ||  :||: :  : |:  :      |:  :| |
Db     284 EESALKKESVRGRK--RKLLPENNYVSRILSTMGATSENASKSCDSDQGNSESTDSGFDR 341

Qy     365 -----RRKPKKVRLLSEIINANQVEDSRSDEVHRENAADPCEDDRST----------IP 408
                ::: :: ::: | : :   | |:  |  :|: ||| :  |||          :|
```



RESULT 2
B47328
natural killer cell tumor-recognition protein - mouse
N;Alternate names: cyclophilin-related NK-tumor recognition protein; natural killer-tumo
C;Species: Mus musculus (house mouse)
C;Date: 17-Jul-1994 #sequence_revision 17-Jul-1994 #text_change 05-Nov-1999
C;Accession: B47328; I77662
R;Anderson, S.K.; Gallinger, S.; Roder, J.; Frey, J.; Young, H.A.; Ortaldo, J.R.
Proc. Natl. Acad. Sci. U.S.A. 90, 542-546, 1993
A;Title: A cyclophilin-related protein involved in the function of natural killer cells.
A;Reference number: A47328; MUID:93133824
A;Accession: B47328
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1507 <AND>
A;Cross-references: GB:L04289; NID:g192866
A;Note: authors translated the codon AGT for residue 972 as Arg
R;Rinfret, A.; Anderson, S.K.

Mol. Immunol. 30, 1307-1313, 1993
A;Title: IL-2 regulates the expression of the NK-TR gene via an alternate RNA splicin
A;Reference number: I57820; MUID:94019422
A;Accession: I77662
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 235-237;263-294 <RIN>
A;Cross-references: GB:S65998; NID:g425701; PIDN:AAB28500.1; PID:g425702
C;Genetics:
A;Gene: NK-TR
C;Superfamily: natural killer cell tumor-recognition protein; cyclophilin homology
C;Keywords: alternative splicing; lymphocyte
F;60-230/Domain: cyclophilin homology <CYP>

Query Match 4.1%; Score 225; DB 2; Length 1507;
Best Local Similarity 19.5%; Pred. No. 0.00031;
Matches 239; Conservative 169; Mismatches 459; Indels 356; Gaps 52;



```
Db    937 NSEQDVTKSRKSDPRRG---SEKEEGEASSDSESEVGQSHIK--------AKPP------ 979

Qy    736 DQYIAEAPTEHW--GRKDAKKLTWEQFKATTRNSPAATCGAQFRPGIQAVDLTSTHVMGS 793
Db    980 ----AKPPTSTFLPGSDGAWKSRRPQSSASESESSCSNLG-NIRGEPQKQKHSKDDLKGD 1034

Qy    794 SSNYASRQPVIAPLDRYAERAVNQVHARNFPSTIATMEASKLCDRRNAGQVVLYPKESMP 853
Db   1035 HTKRA-REKSKAKKDK-KHKAPKRKQAFHWQPPLEFGDDEE--EEMNGKQVTQDPKEKRH 1090

Qy    854 AT-----------HLLRMMDPSTLASFPNYGTSSRNQMESQLH--NSQYAHNQYKGSTST 900
Db   1091 VSEKCEAVKDGIPNVEKTCDEGSSPSKPKKGTLEQDPLAEGGHDPSSCPAPLKVEDNTAS 1150

Qy    901 S---------YGSNLNGKIPLTFED--------------------LSRHQLHDLHRPLRP 931
Db   1151 SPPSAQHLEEHGPG-GGEDVLQTDDNMEICTPDRTSPAKGEVVSPLANHRLDSPEVNIIP 1209

Qy    932 -------HPRVGVLGSLLQKEIANWSENCGTQSGYKLGVSTGITS--HQMNRKEHFEALN 982
Db   1210 EQDECMAHPRAG--GE--QESSMSESKTLG-ESGVKQDSSTSVTSPVETSGKKEGAEKSQ 1264

Qy    983 SGMFSAKWNALQ-LGSVS------SSADFLSARNSIAQSWTRG-----KGK-MVHPLDRF 1029
Db   1265 MNL-TDKWKPLQGVGNLSVSTATTSSALDVKALSTVPEVKPQGLRIEIKSKNKVRPGSLF 1323

Qy   1030 --VRQDICITNKNPADFTTISND 1050
Db   1324 DEVRKTARL-NRRPRNQESSSDD 1345


RESULT   3
T00385
KIAA0624 protein - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000
C;Accession: T00385
R;Ishikawa, K.; Nagase, T.; Suyama, M.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.
DNA Res. 5, 169-176, 1998
A;Title: Prediction of the coding sequences of unidentified human genes. X. The complete
A;Reference number: Z14142; MUID:98403880
A;Accession: T00385
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1983 <ISH>
A;Cross-references: EMBL:AB014524; NID:g3327061; PIDN:BAA31599.1; PID:g3327062
C;Genetics:
A;Note: KIAA0624


  Query Match             4.0%;  Score 221;  DB 2;  Length 1983;
  Best Local Similarity   18.6%;  Pred. No. 0.00074;
  Matches  210;  Conservative  171;  Mismatches  423;  Indels  328;  Gaps   53;

Qy     83 HDQ---KKCDEHKASSSPFSVAKFRRWDCSKCLDKLKTSDNGTAPRTLPAKQNGTSDGCS 139
Db    947 HDEVVDVKCHSHSPFRNERGKGKIRHH--ISCIEKLSKTESISVP---------TSDHRS 995

Qy    140 ITFVRSTFVPASVGSQKVS---------PSTQSS---QGKNADRSTLPKSVQEG------ 181
Db    996 L-------IEANQSNSKVSELDTIYCTLPRKSSSFLIHGRQSGSKIMAASLRNGPPPFQI 1048

Qy    182 -------------NDSKCNAPSGKNGAAEANTDSPMKDLQGPAQNYDVAANVSEDNTSVD 228
Db   1049 KNNVEDAMGNYMLNKFSPSSPESANECSKVLSDSA---LEAPEATERMTNVKSSGSTSVR 1105

Qy    229 VGALP-------EVPQITWHIE----------VNGADQPPSTPKLSEVVLKRNEDENGKT 271
Db   1106 KGPLPFLINRAMSCPSGEPHASTGREGRKKPLTSGMDASELTFRAWERIISPVESDSS-- 1163

Qy    272 EETLVAEQCNLTK----------------------------DPNP----MSGKERDQ 296
Db   1164 -----VRDCSLTKRQHQKENFQEYTEKEGKMAASRRSVFALSNEDPLPFCSDLSGKERGK 1218

Qy    297 VAEQCNLTKDPKPVSGQKCEQICNEPCEEVVLKRSSKSKRKTDKKLMKKQQHSKK----- 351
Db   1219 TLHKVK-TTSTFSVSGDE----DNVKCLEVVSIYYTLPRKPSKKFCNLLQQYTQNTNLLI 1273

Qy    352 RTAQADVSDAKLCRRKPKK-----------VRLLSEIINANQVEDSRSDEVHRENAADP 399
Db   1274 ESPQVETETFPNALEKDKQNYSTREQSGTPSCENLKMSVNSDQTLTTENMTAFRLSNRGP 1333

Qy    400 CEDD-------RSTIPVPMEVSMDIPVSNHTVGEDGLKSSKNKTKRKYSDVVDDGSSLMN 452
Db   1334 LAPTLQEMASVEAAVSLPEEESKAREIFSDNLAKTPLGDSENKKER------------- 1379

Qy    453 WLNGKKKRTGSVHHTVAHPAGNLSNKKVTPTASTQHDDENDTENGLDTNMHKTDVCQHVS 512
Db   1380 ---GKKLQSETLHTSLMLQRKNVSEEK-SENCQQSINSSNSGPSSLPA-LSEVNIGNSQT 1434

Qy    513 EISTQRCSSKGKT-----AGLSKGKTHSAASTKYG--GESTRNGQNIHVLSAEDQCQMET 565
Db   1435 RRSSWECTGSGRAIPFTGSGKCPQKDHTSTAVGDGSSGSQPREGRG----DIGTNCQKMT 1490

Qy    566 ENSVLSHSAK----VSPAEHDIQIMSDLHE--------QSLPKK--KKKQKLEVTREKQT 611
Db   1491 -NKTLSHSESQVFALTPALHKLQLGEETQSDEPNLESLQSEPRELPQRSQEANMTESRK- 1548

Qy    612 MIDDIPMDIVELLAKNQHERQLMTETDCSDINRIQSKTTADDDCVIVAAKDGSDY----- 666
Db   1549 -------------AEDEMQKSAWDQPSLPEGNK--NKTNLDD---LVKGENRSSVKHRLA 1590

Qy    667 ----ASSVFDTN--SQQKSLAS---QSTQKELQGHLALTTQE------SPHPQNFQS--- 708
Db   1591 AMSKASRKFPAKDVSPRRHVATIFPQSGSRSGFDHLSLGTVECNPLFPEPTPKSAESIGE 1650

Qy    709 --TQEQQTHLRMEE---------------MVTIAASSPLFSHHDDQY--IAEAPTEHWGR 749
Db   1651 SRLSENRKHVKKSENLLPITVLPNREPSTHVSNQKSNSISQRHQNEFKNVSESPSKHENS 1710

Qy    750 KDAKKLTWEQFKATTRNSPAATCGAQFRPGIQAVDLTSTHVMGSSSNYASRQPVIAPLDR 809
Db   1711 KD----------VTAAQNLVRESGAP-----SPITFTSLREAEFSDNQRRLSPPF-PLEP 1754

Qy    810 YAERAVNQVHARNFPSTIATMEASKLCDRRNAGQVVLYPKESMPATHLLRMMDPSTLASF 869
Db   1755 -AQK----------SRVSSPLASFLQQQRSASSL-----EWEPEPHLYR---SKSLKSI 1794

Qy    870 PNYGTSSRNQMESQLHNSQYAHNQYKGSTSTSYGSNLNGKIPLTFEDLSRHQLHDLHRPL 929
Db   1795 NVHGDLLRKS-----HPPKVRERHFSESTSI---DNALSRLTLGNEFSVNNGYSRRFRSF 1846

Qy    930 RPHPRVGVLGSLLQKEIANWSENCGTQSGYKLGVSTGITSHQMNRKEHFEALNSGMFSAK 989
Db   1847 SELPSCD--GN------ESWAYRSGTKTGPRSAISI-----------YRPIDYGIFGKE 1886

Qy    990 WNALQLGSVSSSADFLSARNSIAQSWTRGKGKMVHPLDRFVRQDICITNKNP 1041
Db   1887 QQLAFLENVKRS---LTQGRLWKPSFLKNPG--------FLKDDL----RNP 1923


RESULT   4
S62982
vacuolar protein VAC7 - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein N2467; protein YNL054w
C;Species: Saccharomyces cerevisiae
C;Date: 27-Apr-1996 #sequence_revision 03-May-1996 #text_change 07-May-1999
C;Accession: S62982; S58722; S68168
R;Bergez, P.; Doignon, F.; Crouzet, M.
submitted to the Protein Sequence Database, April 1996
A;Reference number: S62975
A;Accession: S62982
A;Molecule type: DNA
A;Residues: 1-1165 <BER>
A;Cross-references: EMBL:Z71330; NID:g1301918; PID:e239893; PID:g1301921; MIPS:YNL054
A;Experimental source: strain S288C
```

```
R;Bergez, P.; Doignon, F.; Crouzet, M.
Yeast 11, 967-974, 1995
A;Title: The sequence of a 44 420 bp fragment located on the left arm of chromosome XIV
A;Reference number: S58711; MUID:96021608
A;Accession: S58722
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1001 <BEW>
A;Cross-references: EMBL:U12141
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1994
R;Bergez, P.; Doignon, F.; Crouzet, M.
Yeast 12, 297, 1996
A;Title: Corrigendum to: the sequence of a 44 458 bp fragment located on the left arm of
A;Reference number: S68168; MUID:97060022
A;Accession: S68168
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1165 <BEF>
A;Cross-references: EMBL:U12141; NID:g1314216; PID:g1098487
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1994
C;Genetics:
A;Gene: SGD:VAC7
A;Cross-references: SGD:S0004999; MIPS:YNL054w
A;Map position: 14L
C;Function:
A;Description: required for normal vacuole inheritance and morphology
C;Keywords: transmembrane protein; yeast vacuole
F;925-941/Domain: transmembrane #status predicted <TMM>


  Query Match             3.9%;  Score 217;  DB 2;  Length 1165;
  Best Local Similarity   18.7%;  Pred. No. 0.0005;
  Matches  204;  Conservative  156;  Mismatches  416;  Indels  314;  Gaps   46;

Qy    155 QKVSPSTQSSQGKNADRSTLPKSVQEGNDSKCNAPSGKNGAAEANTDSPM--KDLQGPAQ 212
Db      6 RKLTVETETVEAPVANNLLL------SNNSNVVAPNPSTPSASTST-SPLHREIVDDSVA 58

Qy    213 NYDVAANVSE------DNTSVDVGALPEVPQITWHIEVNGADQPPSTPKLSEVVLKRNED 266
Db     59 TANTTSNVVQHNLPTIDNNLMDSDATSH-NQDHWHSDINGAGTSMST---SDIPTDLHLE 114

Qy    267 ENGKTEETLVAEQCNLTKDPNPMSGKERDQVAEQCNLTKDPKPVSGQKCEQICNEPCEEV 326
Db    115 HIGSVSST-NNNSNNALINHNPLSSH---------------------------------- 139

Qy    327 VLKRSSKSKRKTDKKLMKKQQHSKKRTAQADVSDAKLCRGKPKKVRLLSEIINANQVEDS 386
Db    140 -LSNPSSSLRNKKSSLLVASN-------PAFASDVELSKGKPAVISNNMPTSNIALYQTA 191

Qy    387 RSDEVHRENAADPCEDDR------------------STDPVPMEVSMDIPVSNHTVGED 427
Db    192 RSANIHGPSSTSASKAFRKASAFSNNTAPSTSNNIGSNTGPAPL-----LPLPS------ 240

Qy    428 GLKSSKNKTK---RKYSDVVDDGSSLM--NWLNGKKKRTGSVHHTVAHPAGNLSNKKV-T 481
Db    241 --LSQQNKPKIIERPTMHVTNSREILLGENLLDDTKAKNAPANST-THDNGPVANDGLRI 297

Qy    482 PTASTQHDDENDTENGLDTNMH-----KTDVCQHVSEISTQRCSS----------KGKTA 526
Db    298 PNHSNADDNENNNKMKKNKNINSGKNERNDDTSKICTTSTKTAPSTAPLGSTDNTQALTA 357

Qy    527 GLSKGKTHSAASTKYGGESTRNGQNIHVLSAEDQCQMETENSVLSHSAKVSPAEHDIQIM 586
Db    358 SVSSSNADNHNNNKKKTSSNNNGNNSNSASNKTNADIKNGNADLSASTSNNNAIND---- 413

Qy    587 SDLHEQSLPKKKKKQ-----------KLEVTREKQTMIDDIPMDIVELL-------AKNQ 628
Db    414 -DSHESNSEKPTKADFFAARLATAVGENEISDSEETFVYESAANSTKNLIFPDSSSQQQQ 472

Qy    629 HERQLMTETDCSDINRIQSKTTA---DDDCVIVAAKDGSDYAS---------SVFDTNSQ 676
Db    473 QQQQPPKQQQQQQNHGITSKISAPLLNNNKKLLSRLKNSRHISTGAILNNTIATISTNPN 532

Qy    677 QKSLASQSTQKELQGH-----LALTTQESPHPQNFQSTQEQQTHLRMEEMVTIAASSPLF 731
Db    533 LNSNVMQNNNNLMSGHNHLDELSSIKQEPPH-----QLQQQQPPMDVQSVDSYTSDNP-- 585

Qy    732 SHHDDQYIAEAPTEHWGRKDAKKLTWEQFKATTRNSPAATC------GAQFRPGIQAVDL 785
Db    586 ---DSNVIAKSPDKRSSLVSLSKVSPHLLSSTSSNGNTISCPNVATNSQELEPN---NDI 639

Qy    786 TSTHVMGSS----------SNYASRQ--------------PVIAPLDRYA--------ER 813
Db    640 STKKSLSNSTLRHSSANRNSNYGDNKRPLRTTVSKIFDSNPNGAPLRRYSGVPDHVNLED 699

Qy    814 AVNQVHARNFPSTIATMEASKLCDRRNAGQVVLYPKESMPATHLLRMMDPSTLASFPNYG 873
Db    700 YIEQPH--NYPTMQNSVKKDEFYNSRN---------NKFP--HGLNFYGDNNVIEEENNG 746

Qy    874 TSS---------------------------------RNQMESQLHNSQYAH-------- 891
Db    747 DSSNVNRPQHTNLQHEFIPEDNESDENDIHSMFYYNHKNDLETKPLISDYGEDEDVDDYD 806

Qy    892 ------NQYKGSTSTSYGSNLNGKIPL----TFEDL-------SRHQLHDLHRPLRPHPR 934
Db    807 RPNATFNSYYGSASNTHELPLHGRMPSRSNNDYYDFMVGNNTGNNNQLNE-YTPLRMKRG 865

Qy    935 VGVLGSLLQKEIANWSENCG-----TQSGYKLGVSTGITSHQM-NRKEHFEALNSGMFSA 988
Db    866 QRHL-SRTNNSIMNGSIHMNGNDDVTHSNINNNDIVGYSPHNFYSRKSPFVKVK------ 918

Qy    989 KWNALQLGSVSSS-------ADFLSARNSIAQSWTRGKGKMVHPLDRFVRQ------DIC 1035
Db    919 --NFLYLAFVISSLLMTGFILGFLLATNKELQDVD------VVVMDNVISSSDELIFDIT 970

Qy   1036 ITNKNPADFT 1045
Db    971 VSAFNPGFFS 980


RESULT   5
T34513
hypothetical protein ZK783.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T34513
R;Favello, A.; Vaudin, M.
submitted to the EMBL Data Library, August 1994
A;Description: The sequence of C. elegans cosmid ZK783.
A;Reference number: Z21536
A;Accession: T34513
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3507 <FAV>
A;Cross-references: EMBL:U13646; PIDN:AAC24418.1; GSPDB:GN00021; CESP:ZK783.1
A;Experimental source: strain Bristol N2; clone ZK783
C;Genetics:
A;Gene: CESP:ZK783.1
A;Map position: 3
A;Introns: 14/1; 48/2; 84/1; 196/3; 303/1; 381/1; 586/1; 605/1; 1175/3; 1207/1; 1409/
 3504/1


  Query Match             3.9%;  Score 215.5;  DB 2;  Length 3507;
  Best Local Similarity   17.9%;  Pred. No. 0.0031;
  Matches  242;  Conservative  179;  Mismatches  510;  Indels  419;  Gaps   52;

Qy     12 RVVGTNCMLARGGTGAVAPVLELTATPRQDAAA--------------EAGVDEPAQHQC- 56
Db   1780 RMGSCGCKCMAGYTGDGATCIKIEEEPKSDKTACTDEWSRLCELEKKQCTVDEEEVPQCG 1839

Qy     57 ----EHFSIRGYVALLQ--------------------KKDPKFCSLSRIF---------- 82
Db   1840 ACLPGHHPINGTCQSLQISGLCAQKNDCNKHAECIDIHPDSHFCSCPDGFIGDGMICDDV 1899
```

```
Qy      83 ----------HDQKKCDEHKASSSPFSVAKFRRWDCSKCL---------DKLK------- 116
Db    1900 DECNNAGMCDDENTKCENTIGSFNCVCLEGFKKVD-EKCVVDEKKQPNREKIEIDEENSS 1958

Qy     117 --------------TSDNGTAPRTLPAKQNGTSDGCSITFVR---STFVPASVGSQKVS 158
Db    1959 SSNSGQEKPTTKGIVSSTSATSSESTTAEPHVTTSISSTTSTKDMTSSKSPENVTMSSES 2018

Qy     159 P--STQSSQGKNADRSTLPKSVQEGNDSKCNAPSGKNGAAEANTDSPMKDLQGPAQNYDV 216
Db    2019 PEVSTSSSKSTTASETTVSSTPSESSSSEAPLTSSPATTTEVITESSVKS---------- 2068

Qy     217 AANVSEDNTSVDVGALPEVPQITWHIEVNGADQPPSTPKLSEVVLKRNEDENGKTEETLV 276
Db    2069 TTPKEESSSEITVKLSSKSPEVT----ESSVKSSPSTPSTTSQSVTSTVPETSKS--TVL 2122

Qy     277 AEQCNLTK---------------DPNPMSGKERDQVAEQCNL-----TKDPK------- 308
Db    2123 SSEAPVTSTSPTEVHTSSETKPSLSASSTTGDTNSTTPSTSSLASVKSTSAPEGTSASVA 2182

Qy     309 PVSGQKCEQICNEP-------CEEVVLKRSSKSKRKTDKKLMKKQQHSKKRTAQADVSDA 361
Db    2183 PVKLSSLSPDVSQPSTKTFDATESSTVQASETSSGTSVKSTSEPESHVTKLSITSSNPSS 2242

Qy     362 KLCRRKPKKVRLL----------------------SEIINANQ---VEDSRSDEVHREN 395
Db    2243 SVPVTSPKSTPTVPESTEQPTSTTPSGQSLTPMNSNSEVLTTSEPHVLSSSLSPDVSQSS 2302

Qy     396 AADPCEDDRSTIPVPMEVSMDIPVSNH---------TVGEDGLKSSKNKTKRKYSDVVDD 446
Db    2303 TTPNNLSESSTVETP-KTSSEVSLNSEEPSTTEAPTTLSPDILSTTTNNLSQSSTVSTED 2361

Qy     447 GSSLMNWLNGKKKRTGS---VHHTVAHPAGNLSNKKVTPTASTQHDD--ENDTENGLDTN 501
Db    2362 RSEISS--ENSEKPTSAPELVTSSVTHVA---SSSPDVPTESSEPDDLTGSSTEN----- 2411

Qy     502 MHKTDVCQHVSEISTQRCSSKGKTAGLSKGKTHSAASTKYGGESTRNGQNIHVLSAEDQC 561
Db    2412 ---------IPEASSKQTISSTPTPDTTTASEEPTKSTSMSPD---------------- 2445

Qy     562 QMETENSVLSHS---------AKVSPAEHDIQIMSDLHEQSLPKKKKKQKLEVTREKQTM 612
Db    2446 -LSTTSNVLSESSTTPESSSKSPVSSSTEGISVVTSTEFSKVPESTISSVLEEDLTKTT- 2503

Qy     613 IDDIPMDIVELLAKNQHERQLMTETDCSDINRIQSKTTADDDCVIVAAKDGSDYASSVFD 672
Db    2504 ----PSPILEETTTASETSEPLTEDSLTVSVRIHELTTSSEN---VPKESESTTTSSESS 2556

Qy     673 TNSQQ------------KSLASQSTQKEL----------QGHLALTT------QESPHPQ 704
Db    2557 KPSQEPAGILTSTVVVPTSSVSLITASEIEAITSNTPFKQGRTPITTSPKSLVKSTTSPS 2616

Qy     705 NFQSTQEQQTHLRMEEMVTIAASSPLFSHHDDQYIAEAPTEHWGRKDAKKLTWEQFK--- 761
Db    2617 TVTSSEPSESTKRTTVSTTVSTTT-------------PTEETTTSESLILTAAPSKPTE 2662

Qy     762 ATTRNSPAATCGAQ------------FRPGIQAVDLTSTHVMGS-----SSNYASRQPVI 804
Db    2663 STTESSEAPTTPAKTSETKPSNVSSTSRKSTENVE-TSTSQSGSLESSTMSSTSSEPETN 2721

Qy     805 APLDRYAERAVNQVHARN-----------------FPSTIATMEASKLCDR-------- 838
Db    2722 APAVTVSSEASSTTLEENSSTSSPTSSEASVKLSSLFPESI-TSEAVTVSSRAPAEITMS 2780

Qy     839 ----RNAGQVVLYPKESMPATHLLRMMDPS--TLASFPNYGTSSRNQMESQLHNSQY--- 889
Db    2781 SESHREISTVSSEPSE--PEIPLSTTVSPNVVTASSIP-----SEEPILSSVTSSSTPRV 2833

Qy     890 -----------------AH-NQYKGSTSTSYGSNLNGKIPLTFEDLSRHQ---------- 921
Db    2834 RLITGTPDDLIVSVTVPSHGNRRQNITASSVPSNSTSPIILPSESLTTPQPPPTTTTTAK 2893

Qy     922 ---------------LHDLHRPLRPHPRVGVLGSLLQKEIANWSENCGTQSGYKLGVST 965
Db    2894 PATTSGKRGPPSIQPPAEMFTTPAPPPPSNGGYGEETNQEEEQVTSTTTTEAP---SLCS 2950

Qy     966 GITSHQMNRKEHFEALNSGMFSAKWNALQLG----SVSSSADFLS-----ARNSIAQSWT 1016
Db    2951 TVTCHSLATCEQ----STGVCICRDGFIGDGTTACSKKSTADCISLPSLCADKAKCDNST 3006

Qy    1017 R---------GKGKMV--HPLDRFVRQDIC 1035
Db    3007 RSCECDAGYIGDGYVCSPHPQDCVLRDNLC 3036


RESULT   6
T13030
microtubule binding protein D-CLIP-190 - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C;Accession: T13030
R;Lantz, V.A.; Miller, K.G.
J. Cell Biol. 140, 897-910, 1998
A;Title: A class VI unconventional myosin is associated with a homologue of a microtu
A;Reference number: Z17588; MUID:98139549
A;Accession: T13030
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1690 <LAN>
A;Cross-references: EMBL:AF041382; NID:g2773362; PID:g2773363; PIDN:AAB96783.1
A;Experimental source: strain Oregon R
C;Genetics:
A;Cross-references: FlyBase:FBgn0020503
C;Keywords: cytoskeleton


  Query Match             3.9%;  Score 215;  DB 2;  Length 1690;
  Best Local Similarity   19.1%;  Pred. No. 0.0013;
  Matches  171;  Conservative  139;  Mismatches  352;  Indels  234;  Gaps   33;

Qy     174 LPKSVQE------------GNDSKCNAPSGKNGAAEANTDSPM---KDLQGPAQNYDVAA 218
Db     401 LQKNINELKARIVELESALGNERK------KTEELQCSIDEAQFCGDELNAQSQVYKEKI 454

Qy     219 NVSEDNTSVDVGALPEVPQITWHIEVNGADQPPSTPKLSEVVLKRNEDENGKTEETLVAE 278
Db     455 HDLESKITKLVSATPSLQSIL----------PPDLP----------SDDGALQEEIAQL 493

Qy     279 QCNLTKDPNPMSGKERDQVAEQCNLTKDPKPVSGQKCEQIC---------NEPCEEVVLK 329
Db     494 QEKMTIQQKEVESRIAEQLEEEQRLRENVKYLN----EQIATLQSELVSKDEALEKFSLS 549

Qy     330 RSSKSKRKTDKKLMKKQQHSKKRTAQADVSDAKLCRRKPKKVRLLSEIINANQVEDSRSD 389
Db     550 ECGIENLRRELALLKEENEKQAQEAQAEFT-RKLAEKSVEVLRLSSELQNLKATSDSLES 608

Qy     390 EVHRENAADPCEDDRSTIPVPMEVSM-DIPVSNHTVGEDGLKSSKNKTKRKYSDVVDDGS 448
Db     609 E--RVNKSDECE------ILQTEVRMRDEQIRELNQQLDEVTTQLNVQKA-------DSS 653

Qy     449 SLMNWLNGKKKRTGSVHHTVAHPAGNLSNKKVTPTASTQHDDENDTENGLDTNMHKTDVC 508
Db     654 ALDDMLRLQKEGT--------------EEKSTLLEKTEKELVQIKEQAAKTLQDKEQLE 698

Qy     509 QHVSEISTQRCSSKGKTAGLSKGKTHSAAS-TKYGGESTRNGQNIHVLSAEDQCQMETEN 567
Db     699 KQISDLKQLAEQEK-----LVREKTENAINQIQLEKESIEQQLALKQNELEDFQKKQSES 753

Qy     568 SVLSHSAKVSPAEHDIQIM----------SDLHEQSLPKKKKKQKL-EVTREKQTMIDDI 616
Db     754 EVHLQEIKAQNTQKDLELVESGESLKKLQQQLEEKTLGHEKLQAALEELKKEKETIIKEK 813

Qy     617 PMDIVELLAKNQHERQLMTETDCSDINRIQSKTTADDDCVIVAAKDGSDYASSVFDTNSQ 676
Db     814 EQELQQLQSKSAESESALKVVQV-QLEQLQQQAAA-------SGEEGSKTVAKLHDEISQ 865
```

```
Qy     677 QKSLASQSTQKELQGHLALTTQESPHPQNFQ----------------STQEQQTHLRMEE 720
           || | : || ||:     :|: :   :: |                  ||| | |: |
Db     866 LKSQA-EETQSELK-----STESNLEAKSKQLEAANGSLEEEAKKSGQLQEQITKLKSEV 919

Qy     721 MVTIAASSPLFSHHDDQYIAEAPTEHWGRKDAKKLTWEQFKATTRNSPAATCGAQFRPGI 780
           | ||   | |:| |                        :: |:   ||    :
Db     920 EETQAA---LSSYHTD----------------------VESKTKQLEAANAALE----- 948

Qy     781 QAVDLTSTHVMGSSSNYASRQPVIAPLDRYAERAVNQVFA-----RNFPSTIATMEASKL 835
                   :  ||  :   : |   :   : :||     |:  | : | : ||
Db     949 ----------KVNKEYAESRAEASDLQDKVKEITDTLFAELQAERSSSSALHT-KLSKF 996

Qy     836 CDRRNAGQVVLYPKESMPATHLLRMMDPSTLASFPNYGTSSRNQMESQLHNSQYAHNQYK 895
           |    |   | |    :  :|:                ::  || :|| :  : |
Db     997 SDEIATGHKELTSKADAWSQEMLQK----------EKELQELRQQLQDSQDSQTKLK 1043

Qy     896 GSTSTSYGSNLNGKIPLTFEDLSRHQLHDLHRPLRPHPFVGVLGSLLQKEIA-----NWS 950
                 |         ||:  ::                    ||:|:      |
Db    1044 AEGERKEKS---------FEESIKN--------------------LQEEVTKAKTENLE 1073

Qy     951 ENCGTQSGYK-LGVSTGITSHQMNRKEHF---EALNSGMFSAKWNALQLGSVSSSA 1002
           : |||:  | |     ||: ::  ||     :|          |:|: : : ||
Db    1074 LSTGTQTTIKDLQERLEITNAELQHKEKMASEDAQKIADLKTLVEAIQVANANISA 1129


RESULT   7
T42727
proliferation potential-related protein - mouse
C;Species: Mus musculus (house mouse)
C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 02-Sep-2000
C;Accession: T42727
R;Witte, M.M.; Scott, R.E.
submitted to the EMBL Data Library, November 1998
A;Reference number: Z22246
A;Accession: T42727
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1560 <WIT>
A;Cross-references: EMBL:U83913; NID:g3858884; PID:g3858885; PIDN:AAC72432.1
A;Experimental source: strain Balb/C
C;Genetics:
A;Gene: P2P-R
C;Function:
A;Description: involved in hnRNP association and Rb1 binding
C;Superfamily: RING finger homology
F;57-107/Domain: RING finger homology <RRN>


  Query Match             3.9%;  Score 213.5;  DB 2;  Length 1560;
  Best Local Similarity   18.9%;  Pred. No. 0.0014;
  Matches  181;  Conservative  150;  Mismatches  374;  Indels  251;  Gaps   38;

Qy      84 DQKKCDEHKASSSPFSVAKFRRWDCSKCLDKLKTSDNGIAPRTL---PAKQNGT------ 134
           :::| :| ::  :|  : |||:   |  :|: ||        ||   |:::: |
Db     709 NEEKGEESESFLNPELLGKFRKCRGSSGIDETKTD-------TLFVFPSREDATPVRDEP 761

Qy     135 SDGCSITFVRSTFVPASVGSQKVSPSTQSSQGKNADRSILPKSVQEGNDSKCNAPSGKNG 194
           |  ||||                :|   |:      ||   :    |        :|
Db     762 MDAESITF----------------KSVSDKDKREKDKPKVKSDKTKRK------SDG 796

Qy     195 AAEANTDSPMKDLQGPAQNYDVAANVSEDNTSVDVGALEEVPQITWHIEVNGADQPPSTP 254
           :| |  |: :|  :|| :  |               |   : |:           :||
Db     797 SATAKKDNVLKPSKGPQEKVD-------------GDREKSPR----------SEPP--- 829

Qy     255 KLSEVVLKRNEDENGKTEETLVAEQCNLTKDPNPMSGKERDQVAEQCNLTKDPKPVSGQK 314
                ||: ::|  | :   |    :  ||   ::|  |   ::     :: ||   :|
Db     830 ------LKKAKEEATKIDS--VKPSSSSQKD-EKVTGTERKAHSKSAKEHQEAKPAKDEK 880

Qy     315 CEQICNEPCE----------------EVVLKRSSKSKFKTDKKLMKKQ---------- 346
           ::  |::  :               :::  :  | ||||::| : |
Db     881 VKKDCSKDIKSEKPASKDEKAKKPEKNKLLDSKGEKRKFKTEEKSVDKDFESSSMKISKV 940

Qy     347 ---:---QHSKKRTAQADVSDAKLCRRKPKKVRLLSEIINANQVEDSRSDEV---HRENAA 397
                :  | ||  : ||   ||  | |:| :: |    | ::: :|       : |||:
Db     941 EGTEIVKPSPKRKMEGDVE--KL-ERTPEKDKIASSTTPAKKIKLNRETGKKIGNAENAS 997

Qy     398 ---DPCEDDRSTIPVPMEVSMDIPVSNHTVGEDGLKSSKNKTKRKYSDVVDDGSSLMNWL 454
              :| |   ||      :  :          | :|  |    |   |:      | | :
Db     998 TTKEPSEKLESTSSKIKQEKVKGKAKRKVAGSEGSSS----TLVDYTSTSSTGGSPVRKS 1053

Qy     455 NGKKKRTGSVHHTVAH-------PAGNLSNKKVTPTASTQHDDENDTENGLDTNMHKTDV 507
              |   | :|  |:          || ::      | :    ||    |  : |
Db    1054 EEKTDTTRTVIKTMEEYNNDNTAPAEDVIIMIHVPQSKWDKDDFESEEEDVKT------- 1106

Qy     508 CQHVSEISTQRCSSKGKTAGLSKG-KTHSAASTKYGGESTRNGQNIHVLSAEDQCQMETE 566
                 ||   | || : : |   |  :|: ||  : :   :  :  |       |
Db    1107 --------TQPIQSVGKPSSIIKNVTTKPSATAKYTEKESEQPEKLQKL------PKEAS 1152

Qy     567 NSVLSH---SAKVSPAEHDIQIMSDLHEQSLPKKKKKQKLEVTREKQTMIDDIPMDIVEL 623
           : :: |   |:| | :     :      |  |       |:|     :|:: :|
Db    1153 HELMQHELRSSKGSASSEKGRAKDREHSGSEKDNPDKRKSGAQPDKESTVD--------R 1204

Qy     624 LAKNQHERQLMTETDCSDINRIQSKTTADDDCVIVAA-KDGSDYASSVFDTNSQQKSLAS 682
           |::  | : |   :         :::|:   : |  :: || :        | :|:     |
Db    1205 LSEQGHFKTLSQSSK-------ETRTSEKHESVRGSSNKDFTPGRDKKVDYDSRD---YS 1254

Qy     683 QSTQKELQGHLALTTQESPHPQNFQSTQEQQTHLRMEEMVTIAASSPLFSH--------- 733
            | ::: :| ||   ::|| |:  :|   |:: || |  :    :    |:
Db    1255 SSKRRDERGELA-RRKDSP-PRGKESLSGQKSKLREERDLPKKGAESKKSNSSPPRDKKP 1312

Qy     734 --HDDQYIAEAPTEHWGRKDAKKLTW---EQFKATTRNSPAATCGAQFRPGIQAVDLTST 788
             |   |  : | |     |   |::     |:  |  ||
Db    1313 HDHKAPYETKRPCEE--TKPVDKISGKEREKHAAEARN---------------------- 1348

Qy     789 HVMGSSSNYASRQPVIAPLDRYAER--AVNQVHARNFPSTIATMEASKLCD--RRNAGQV 844
              |  |:       ::  | |    |:  || ||              :|:|     |   :
Db    1349 ---GKDSSGGKLPCILNPPDLPMEKELAVGQVEKSAVKPKPQLSHSSRLSSDLTRETDEA 1405

Qy     845 VLYP------------KESMPATHLLRMMDPST----------LASFP----NYGTSS 876
              |              ||      : : :   |           || |    :   ||
Db    1406 AFEPDYNESDSESNVSVKEEEAVASISKDLKEKTTEKAKESLTVATASQPGADRSQSQSS 1465

Qy     877 RNQMESQLHNSQYAHNQYKGSTSTSYGSNLNGKIPLTFEDLSRHQLHDLHRPLRPH 932
           :   |: |:   :  :   |:::| ||  : |     ::  :|: |  |:  : |
Db    1466 PSVSPSRSHSPSGSQTRSHSSSASSAGSQ-DSKKKKKKKEKKKHKKHKKHKKHKKH 1520


RESULT   8
T06310
hypothetical protein F11C18.80 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 22-Oct-1999
C;Accession: T06310
R;Bevan, M.; Terryn, N.; Ardiles, W.; Buysshaert, C.; Dasseville, R.; De Clerck, R.;
ewes, H.W.; Mayer, K.F.X.; Schueller, C.
submitted to the Protein Sequence Database, April 1999
A;Reference number: Z15589
A;Accession: T06310
A;Molecule type: DNA
A;Residues: 1-852 <BEV>
A;Cross-references: EMBL:AL049607; GSPDB:GN00062; ATSP:F11C18.80
A;Experimental source: cultivar Columbia; BAC clone F11C18
C;Genetics:
A;Gene: ATSP:F11C18.80
A;Map position: 4
A;Introns: 34/3; 102/3; 163/2; 205/3; 269/3; 282/3; 564/3; 602/3; 622/2; 639/3; 667/3


  Query Match             3.8%;  Score 209.5;  DB 2;  Length 852;
  Best Local Similarity   21.2%;  Pred. No. 0.001;
  Matches  174;  Conservative   97;  Mismatches  282;  Indels  269;  Gaps   39;

Qy      31 VLELTATPRQ-----DAAAEAGVDEPAQHQCEHFSIRGY-------------VALLQKK 71
           :||  |  |       |   :| : |  ||  :  :|| :             : | : :
Db     129 ILETVAKVRSCVVMLDLECDALLIEMFQHFLK--AIRDHHSGNVFSSMENIMTLVLEESE 186
```

```
Qy      72 DPKFCSLSRIFHDQKKCDEHKASSSPFSVAKFRRWDCSKCLDKLKT------SDNGTAPR 125
Db     187 DIPSEMLSPILHSVKKDDEISQVS-----RRLAEQVLSNCASKLKTYLTEAVKSSGV--- 238

Qy     126 TLPAKQNGTSDGCSITFVRSTFVPASVGSQKVSPSTQSSQGKNADRSTLPKSVQEGNDSK 185
Db     239 PLDKYSNIVASICEGTF------SALQQDQVVANEKEDSQGHIKRETEVEKAAEISTPER 292

Qy     186 CNAP---SGK----NGAAEAN-----TDSPMKD----LQGPAQNYDVAANVSEDNTSVDV 229
Db     293 TDAPKDESGKSGVSNGVAQQNDSSVDTDSMKKQDDTGAKDEPQQLDNPRNTDLNNTT--- 349

Qy     230 GALPEVPQITWHIE--------VNGAD-----------------------QPPSTPKLS 257
Db     350 ---EEKPDVEHQIEEKENESSSVKQADLSKDSDIKEETEPAELLDSKDVLTSPPVDSSVT 406

Qy     258 EVVLKRNE--------------DENGKTEETLVAEQCNLTKDPNPMSGKERDQVAEQCNL 303
Db     407 AATSSENEKNKSVQILPSKTSGDETANVSSPSMAEELPEQSVPKKTANQKKKE-----SS 461

Qy     304 TKDPKPVSGQKCEQICNEP--CEEVVLKRSSKSKRKTDKKLMKKQQHSKKRTAQADVSDA 361
Db     462 TEEVKPSASIATEEVSEEPNTSEPQVTKKSGK-KVASSSKTKPTVPPSKKST-----SET 515

Qy     362 KLCRRKPKKVRLLSEIINANQVEDSRSDEVHRENAADPCEDDRSTIPVPMEVSMDIPVSN 421
Db     516 KVAKQSEKKV-----VGSDNAQESTKPKEEKKKPGRGKAIDEESL--------------- 555

Qy     422 HTVGED--------GLKSSKNK--------------TKRKYS---------DVVDDGSSL 450
Db     556 HTSSGDNEKPAVSSGKLASKSKKEAKQTVEESPNSNTKRKRSLGQGKASGESLV--GSRI 613

Qy     451 MNW--------------LNGKKKRTGSVHHTVAHPAGN-----LSNKKVTPTASTQ--HD 489
Db     614 KVWWPMDQAYYKGVVESYDAAKKK-----HLVIYDDGDQEILYLKNQKWSPLDESELSQD 668

Qy     490 DENDTENGLDTNMHKTDVCQHVSEISTQRCSSKGKTAGLSKGKTHS-----AASTKYGGE 544
Db     669 EEAADQTGQE---------EDASTVGSGAGSSKAKATPASKSSKTSQDDKTASKSKDSKE 719

Qy     545 STRNGQNIHVLSAEDQCQMETENSVLSHSAKVSPAEHDIQIMSDLHEQSLPKKKKKQKLE 604
Db     720 ASREEE----ASSEEESEEEEPPKTVGKSGS-SRSKKDISSVSKSGKSKASSKKKEEPSK 774

Qy     605 VT---REKQTMIDDIPMDIVELLAKNQHERQLMTETDCSDINRIQSKTTADDDCVIVAAK 661
Db     775 ATTSSKSKSGPVKSVP-------------------------AKSKTGKG------KAK 801

Qy     662 DGSDYASSVFDTNSQQKSLASQSTQKELQGHLALTTQESPHP 703
Db     802 SGS--AST---PASKAKESASESESEE--------TPKEPEP 830
```

RESULT 9
T24587
hypothetical protein T06E4.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T24587
R;Lloyd, C.
submitted to the EMBL Data Library, April 1996
A;Reference number: Z19910
A;Accession: T24587
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1295 <WIL>
A;Cross-references: EMBL:Z70756; PIDN:CAA94789.1; GSPDB:GN00023; CESP:T06E4.1
A;Experimental source: clone T06E4
C;Genetics:
A;Gene: CESP:T06E4.1
A;Map position: 5
A;Introns: 17/3; 218/2; 385/3; 481/3; 946/3; 1034/3; 1222/2

Query Match 3.8%; Score 208.5; DB 2; Length 1295;
Best Local Similarity 18.4%; Pred. No. 0.002;
Matches 208; Conservative 179; Mismatches 402; Indels 339; Gaps 52;

```
Qy      84 DQKKCDEHKASSSPFSVAKFRRWDCSKCLDKLKT-SDNGTAPRTLPAKQNGTSDGCSITF 142
Db      96 DNSVCEPPRASSA---------HDVDPNLEQTKVYSEDSISKDCSIADESFATDG-DVTL 145

Qy     143 VRSTFVPASVGSQKVSPSTQSSQGKNADRSTLPKSVQEGNDSKCNAPSGKNG----AAEA 198
Db     146 ANDFFQQATIQDLENTPST----------PLRSQTQVGDESTPTAPSPLRSKIPIESEV 194

Qy     199 NTDSPMKDLQGPAQN---------YD---VAANVSEDNTSVD----VGALPEVPQITWHI 242
Db     195 VQSTSVESVQSTSQHEKSRIEYSLYEFEMLEKNIEHDEVVKDLQLKIELLGEKHRQTL-V 253

Qy     243 EVNGA---DQPPSTPKLSEVVLK-RNEDENGKTEETLVAE-----QCNLTKDPNPMSGKE 293
Db     254 EIKEAREVEEKMLMQQVEEAVKKTKSEREAAKTLEQLLEKRIEELEMKLT-EPN----GE 308

Qy     294 RDQVAEQCNLTKDPKPVSGQKCEQICNEPCEEVVLKRSSK--SKRKTDKKLMKKQQHSKK 351
Db     309 KLQFEHQLEELK-------SRCEELTDKALKVDVMQHSIEDYEKKFVELQEMKEEADEQL 361

Qy     352 RTAQADVSDAKLCRRKPKKVRLLS----EIINANQVEDSRS----------DEVHR-ENA 396
Db     362 QKAKEDIETLQM-----KYVELETTINKEVFSNSEIETLKSEHEIVRKLMLDEIHRLENE 416

Qy     397 ADPCEDDRSTIPV-PMEVSMD---IPVSNHTVGEDGLKSS----KNKTKRKYSDVVDDGS 448
Db     417 MSALQPKNDTTELEELQKTLDDLKIDCCNLTSKMLELQSELVEVKEKATSEIGEAVQKNG 476

Qy     449 SLMNWLNGKKKRTGSV---------HHTVAHPAGNLSNKKVTPTASTQHD---------D 490
Db     477 ELLEQINSLRVENAKLVDMEGQLNDAHRKAEDKDVRISELLTTIESLRQDSEASDKLLMD 536

Qy     491 ENDTENGLDTNMHKTDVCQHVSEISTQRCSSKGKTAGLSKGKTHSAASTKYGGESTRNGQ 550
Db     537 SESTQNEYSLALENT-----VSELETMRREYKA---------------------------- 564

Qy     551 NIHVLSAEDQCQMETENSVLSHSAKVSPAEHDIQIMS-DLHEQSLPKKKKKQ--KLEVTR 607
Db     565 -----SVDKVCSLQLELEEIQHETSVELEEAEIRIKELELAQEEAVKTGSSQLKKLEIVQ 619

Qy     608 EK-QTMIDDIPMDIVELL-------------AKNQHERQ------LMTETDCSDIN--- 643
Db     620 EDCQKLRDQLKEEQIQQLVSLRETSEVMHQESARHQEEKYQIQSKLMSTEAEVIELRSSI 679

Qy     644 -------RIQSKTTADDDCVI------------VAAKDGSDYASS---VFDTNSQQKSL- 680
Db     680 DSLQAEVRVQSDSAADQKHILEDYLRKIRQAEETNEKLRSDLASSEEQILDLKNQQESLI 739

Qy     681 --------ASQSTQKELQGHLALTTQESPHPQNFQSTQEQQTHLRMEEMVTIAASSPLFS 732
Db     740 DDLKEKLHSAESTNQELQVSLEMLKIEVSNAR--QKVMESEVLKESFEALQLELSA---S 794

Qy     733 HHDDQYIAEAPTEHWGRKDAKKLTWEQFKATTRNSPAATCGAQFRPGIQAVDLTSTHVMG 792
Db     795 QEVSRSVVDAAVQ---EKDGLLRLVDTLKLKIEDTEKS----------AQDLQQSSV-- 838

Qy     793 SSSNYASRQPVIAPLDRYAERAVNQVHARNFPSTIATMEASKLCDRRNAGQVVLYPKESM 852
Db     839 -----EEIKQLQLDLQNFKQNAEVLESL---------------NEKL 865

Qy     853 PATHLLRMMDPSTLASFPNYGTSSRNQMESQLHNSQYAHNQYKGSTSTSYGSNLNGKIPL 912
Db     866 NSSH---KRDMVALAS----------QLEELQHKLVVGESQVENVKEELIGAKI------ 906

Qy     913 TFEDLSRHQLHDLHRPLRPHPRVGVLGSLLQKEIANWSENCGTQSGYKLGVSTGITSHQM 972
Db     907 ----MNKEMVDELN---------AKLGDALE----------GME---ELKKSLEVSEAKV 940
```

```
Qy    973 NRKEH------------------FEALNSGMFSAKWNALQLGSVSSSADFLSARNSIAQ 1013
Db    941 QRREEELIAQVSKHRDQQEQLQLTLDELKSAQHSTETSRSQSNELAARIEELEASISFAQ 1000

Qy   1014 SWTRGKGKMVHPLDRFVRQDICITNKNPA------DFTT--ISNDNEY 1053
Db   1001 KALQDVEDVKH------QQDIQISEANEAMVKLKQDFETERTSLQNEF 1042
```

RESULT 10
A53577
ascites sialoglycoprotein 1 - rat (fragments)
C;Species: Rattus norvegicus (Norway rat)
C;Date: 12-Apr-1995 #sequence_revision 12-Apr-1995 #text_change 07-Feb-1997
C;Accession: A53577
R;Wu, K.; Fregien, N.; Carraway, K.L.
J. Biol. Chem. 269, 11950-11955, 1994
A;Title: Molecular cloning and sequencing of the mucin subunit of a heterodimeric, bifun
A;Reference number: A53577; MUID:94216302
A;Accession: A53577
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1630 <WUA>
A;Cross-references: GB:U06746
C;Keywords: glycoprotein

Query Match 3.8%; Score 208; DB 2; Length 1630;
Best Local Similarity 20.2%; Pred. No. 0.0029;
Matches 206; Conservative 120; Mismatches 404; Indels 292; Gaps 41;

```
Qy    106 WDCSKCLDKLKTSDNGTAPRTLPAKQNGTSDGCSITFVRSTFVPASVGSQKVSPSTQS-S 164
Db     13 WLCLSCL--------CSCLLLLPVNTSTTSAPKTSTALPSSTNPSQMTSQVSNPTASSYR 64

Qy    165 QGKNADR----------STLPKSVQEGNDSKCNAPSGKNGAAEANT-------------- 200
Db     65 MTKNTGQASPMVTSSSITTLPQSQHTGSMKTTRNPQ-TTSTTEVTTTLSASSSDQVQVET 123

Qy    201 --------------DSPMKDLQGPAQNYDVAANVSEDNTSVDVGALPEV-----PQITWH 241
Db    124 TSQTTLSPDTTTTSHAPRESSSPPSTSVILTTTASTEGTSGDTGHTMAVTTQGSTPATTE 183

Qy    242 IEVNGADQPPSTPKLSEV----------------------VLKRNEDENGKTEETL---V 276
Db    184 ISVT-----PSTQKMSPVSTFSTSTQEITTLSQSQHTGGMKTTRNPQTTGTTEVTTTLSA 238

Qy    277 AEQCNLTKDPNPMSGKERDQVAEQCNLT---------KDPKPVS---GQKCEQICNEPCE 324
Db    239 SSSDHPTSSPESTPGNTAPRTTETSTTTTTKVLMTSLQQKLPTGSTLGTSTQELTTLPQS 298

Qy    325 E--VVLKRSSKSKRKTDKKLM-----------KKQQHSKKRTAQADVSDAKLCRRKPKKV 371
Db    299 QHTGIMKTTSRTQTTTPTEVTTRTLSASSSDHRQAETSSQTTLSPDTTTTSHAPRESSPP 358

Qy    372 RLLSEIINANQVEDSRSDEVH----RENAADPCEDDRSTIPVPMEVSMDIPV-------- 419
Db    359 S-TSVILTHGHREGTSGDTGHTMAVTTQGSTPATTEISVTPSTQKMS---PVSTFSTSTQ 414

Qy    420 -------SNHTVGEDGLKSSKNKTKRKYSDV----VDDGSSLMNWLNGKKKRTGSVHHTV 468
Db    415 EITTLSQSQHT---GGMKTTRNPQRTTPTEVTTSTLSASSSDQVQVETTSRATLSPDTTT 471

Qy    469 AHPAGNLSNKKVTP-------------TASTQHDDENDTENGLDTNMHKTDVCQHVSEIS 515
Db    472 TSHAPSVSSSSPSPPSTEGTSVDTGLTTAVTTQDSTPATTQGSLTSSSQTLSTVSPLSTS 531

Qy    516 TQRCSSKGKTAGLSKGKTHSAASTKYGGESTRNGQNIHVLSA--EDQCQME-TENSVLSH 572
Db    532 TQETSTQELTS--SQSQHTGSMKTTHNPQTTRNTEVTTTLSASSSDQVQVETTSQTTLSD 589

Qy    573 SAKVSPAEHDIQIMSDLHEQSLPKKKKKQKLEVTREKQTHIDDIPMDIVELLAKNQHERQ 632
```

```
Db    590 ATTTSHAP---------RESSSPPSTS------------------DILTTMAS------ 615

Qy    633 LMTETDCSDINRIQSKTTADDDCVIVAAKDGSDYASS-VFDTNSQQK----SLASQSTQK 687
Db    616 --TEGTSGDTGHTMAVTT----------QGSTPATTEISVTPSTQKMSPVSTFSTSTQE 662

Qy    688 ELQGHLALTTQESPHPQNFQSTQEQQTHLRMEEMVTIAASSPLFSHHDDQYIAE------ 741
Db    663 ------ITTLSQSQHTGGMKTTRNPQTTGTTEVTTTLSASS------SDQVQAETSSQTT 710

Qy    742 -----APTEHWGRKDAKKLTWEQFKATTRNSPAATCGAQFRPGIQAVDLTSTHVMGS--- 793
Db    711 LSPDTTTTSHAPRESSSPPSTSDMLTTT----ASTEGTSGDTG----HTTAVTTQGSIPA 762

Qy    794 ----SSNYASRQPVIAPLDRYAERAVNQVHARNFPSTIATMEASKLCDRRNAGQVVLYPK 849
Db    763 TTQLSTTFASQ----------------KMSTVSTPTTSSIQELSTLPQSQHTGSMEI--- 803

Qy    850 ESMPATHLLRMMDPSTLASFPNYGTSSRNQMESQLHNSQYAHNQYKGSTSTS------YG 903
Db    804 SSRPQTTSV----TSTLSSSPSGSTPV----------------QTRSVTSSSDERTNPTS 843

Qy    904 SNLNGKIPLTFEDLSRHQLHDLHRPLRPHPRVGVLGSLLQKEIANWSENCGTQSGYKLGV 963
Db    844 SGVSNTSPATTEVLTPTSSPE-STPGNTAPRTTETSTTTTTKVLMTSLQQKLPTGSTLGT 902

Qy    964 ST 965
Db    903 ST 904
```

RESULT 11
T44231
hypothetical protein U86 [imported] - human herpesvirus 6 (strain Z29)
C;Species: human herpesvirus 6
A;Variety: strain Z29
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 02-Jun-2000
C;Accession: T44231
R;Dominguez, G.; Dambaugh, T.R.; Stamey, F.R.; Dewhurst, S.; Inoue, N.; Pellett, P.E.
J. Virol. 73, 8040-8052, 1999
A;Title: Human herpesvirus 6B genome sequence: coding content and comparison with hum
A;Reference number: Z22734; MUID:99412318
A;Accession: T44231
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1520 <DOM>
A;Cross-references: EMBL:AF157706; PIDN:AAD49674.1
A;Experimental source: strain Z29; variant B
C;Genetics:
A;Note: U86

Query Match 3.8%; Score 207.5; DB 2; Length 1520;
Best Local Similarity 18.6%; Pred. No. 0.0028;
Matches 200; Conservative 170; Mismatches 480; Indels 227; Gaps 41;

```
Qy     84 DQKKCDEHKASSSPFSVAKFRRWDCSKCLDKL-KTSDNGTAPRTLPAK---QNGTSDGCS 139
Db    270 DNKEHKKQQVSTSPDAISLSYR-PSSQKMDLIKKIYDTEVIP--LPKEALIDNGSYCGVD 326

Qy    140 ITFVRSTFVPASVGSQKVSPSTQSSQGKNADRSTLPK-SVQEGNDSKCNAPSGKNGAAEA 198
Db    327 TQKYKKTHIRCR--------SIQKTKGHSS--QTINKHKVQKHNEN--HVPS-------- 366

Qy    199 NTDSPMKDLQGPAQNYDVAANVSEDNTSVD--VGALPEVPQITWHIEVNGADQPPSTPKL 256
Db    367 --RSDLKQRKS-NQHEDEAVTEARDFSKLDPLLSPLPMTPE----------------PTL 407

Qy    257 SEVVLK---RNEDENGKTEETLVAEQCNLTKDPNPMSGKERDQVAEQCNLTK---DPKPV 310
Db    408 NFAVHKTKIHSDSELHHTKKNIHRSKTSL--QDRVLISKHAPRAPTKDNSYKKHHDPKDT 465

Qy    311 SGQKCEQICNEPCEEVVLKRSSKSKRKTDKKLMKKQQHSKKRTAQADVSDAKLCRRKPKK 370
```

```
Db    466 NDPK--------------MKHSRGRTTSKKNTANSNGHQDVREVAVKNVSGKAAKNTSKK 511

Qy    371 VRLLSEIINANQVEDSRSDEVHRENAADPCEDDRSTIPVPMEVSMDIPVSNHTVGEDGLK 430

Db    512 ----SD--PSHNLHGKTSDEQYKTS------PDNEKISTPPKSKTHHCIHDSSSSEEGQY 559

Qy    431 SSKNKTKRKYSDVVDDGSSLMNWLNGKKKRTGSVHHTVAHPAGN-LSNK------KVTPT 483

Db    560 KSPNNSESNY------------------------------GNCLSDKFYDNFTKTTPN 587

Qy    484 ASTQHDDENDTENG-----LDTNMHKTDV-------------------------CQHVS 512

Db    588 SKTNHKTEESTENTDLNSFSNENTNKTEIEDSNIIQPFSQLFCNETIIPSTSACPTQETP 647

Qy    513 EISTQRCSSKGKTAGLSKG-------KTHS------AASTKYGGESTRNGQNIHVLSAED 559

Db    648 STMNRNCASKKHSAGANKNLTDNSPIRSHSNPSSFTAFNKSNSGNSTMNS-----TSNGD 702

Qy    560 QCQMETENSVLSHSAKVSPAEHDIQIMSDLHEQSLPKKKKKQKLEVTREKQTMIDDIPMD 619

Db    703 ECTDKKPN--------CSSTENKSETSNQTNGENSDKPLSKTFTEVSDRASSRASSRASS 754

Qy    620 IVELLAKNQHERQLMTETDCSDINRIQSK--TTADDDCVIVAAKDGSDYASSVFDTNSQQ 677

Db    755 RASSRASSRASSRASSRASSRTSSRASSRAFSRASSRVSSRASSRASSRASSRASSRASS 814

Qy    678 KSLASQSTQKELQGHLALTTQESPHPQNFQSTQ-EQQTHLRMEEMVTIAASSPLFSHHDD 736

Db    815 RASGRASGRASSRASSRVSSRASSRASSRASSRASSRASSRASSRVSSKASSRASSRASS 874

Qy    737 QYIAEAPTEHWGRKDAKKLTWEQFKATTRNSPAATCGAQFRPGIQAVDLTSTHVMGSSSN 796

Db    875 RASSRASSRASSRASSRASSRASSRASSRASSRASSRASSRASSRASSRASSRASS 934

Qy    797 YASRQPVIAPLDRYAERAVNQVHARNFPSTIATMEA-SKLCDRRNAGQVV-------LYP 848

Db    935 RASGRASSKASSRASSRASGKTPNNKLMSNIPSPQTYEKTSNKRKPRQIYCDSNKRQIYP 994

Qy    849 KESMPATHLLRMMDPSTLASFPNYGTSSRNQMESQLHNSQYAHNQY--------KGSTST 900

Db    995 HDTSISTEAEVSEIKFRCGPELNFYKNAAARLQSFNHNDQFYNPRFRPHIRTNRKKSEST 1054

Qy    901 SYGSNLNGKIPLTFEDLSRHQLHDLHRPLRPHPRVGVLGSLLQKEIANWSENCGTQSGYK 960

Db   1055 NYTDSESS--TSRSRSHSRYSPDSLNTPKRKKHKSG--SSSISSSI---EENSRSNSRTE 1107

Qy    961 LGVSTGIT---SHQMNRKEHFEALNSGMFSAKWNALQLGSVSSSAD--FLSARNSIAQ-- 1013

Db   1108 TGTDKLITFNYQHSRTRSSSSSSVSSSCSSSRSKSWRKSRYKNSEDPATYSPRSRLQQLE 1167

Qy   1014 ------SWTRGKGKMVHP--------LDRFVRQDICITNKNPADFTTISNDNEYMDY 1056

Db   1168 KQTRPKSASRDKTKIKSPNHESKHRHADMF-RNSQKTGEKFPLDNSSPSNTHEQSNH 1223


RESULT 12
T13564
microtubule-associated protein homolog - fruit fly (Drosophila melanogaster)
N;Alternate names: hypothetical protein EG:49E4.1
C;Species: Drosophila melanogaster
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C;Accession: T13564
R;Spanos, L.; Papagiannakis, G.; Siden-Kiamos, I.; Louis, C.
submitted to the EMBL Data Library, April 1999
A;Description: Sequencing the distal X chromosome of Drosophila melanogaster.
A;Reference number: Z17689
A;Accession: T13564
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-5327 <SPA>
A;Cross-references: EMBL:AL031128; PIDN:CAA20006.1
C;Genetics:
A;Cross-references: FlyBase:FBgn0025392
A;Introns: 24/2; 52/3; 104/3; 179/1; 232/1; 1669/3; 2566/1; 4798/3; 5272/1
A;Note: EG:49E4.1
C;Superfamily: Drosophila 576K microtubule-associated protein homolog


  Query Match             3.7%;  Score 204.5;  DB 2;  Length 5327;
  Best Local Similarity   19.1%;  Pred. No. 0.021;
  Matches  225;  Conservative  174;  Mismatches  457;  Indels  319;  Gaps   51;

Qy     31 VLELTATPRQDAAAEAGVDEPAQHQCE---HFSIRGYVA----LLQKKDPKFCSLSRIFH 83

Db   1967 VAEKSPLPSKEASRPASVAESIKDEAEKSKEESRRESVAEKSPLPSKEASRPASVAESIK 2026

Qy     84 DQ-KKCDEH--------------KASSSPFSVAKFRRWDCSKCLDKLKTSDNGTAPRTLP 128

Db   2027 DEAEKSKEESRRESVAEKSPLPSKEASRPASVAESIKDEAEKSKEESR-RESAAEKSPLP 2085

Qy    129 AKQNGTSDGCSITFVRSTFVPASVG---SQKVSPSTQSSQGKNADRSTLPKSVQEGNDSK 185

Db   2086 SKEASR--------------PASVAESVKDEADKSKEESRRESMAESGKAQSI-KGDQSP 2130

Qy    186 CNAPSGKNGAAEANTDSPMKDLQGPAQNYDVAANVSEDNTSVDVGALPEVPQITWHIEVN 245

Db   2131 LKEVSRPESVAESVKDDPVKSKE-PSRRESVAGSVTADSARDDQSPL----------ESK 2179

Qy    246 GADQPPS-TPKLSEVVLKRNEDENGKTEETLVAEQCNLTKDPNPMSGKERDQVAEQCNLT 304

Db   2180 GASRPESVVDSVKDEAEKQESRRESKTESVI---------PPKAKDDKSPKEVLQPVSMT 2230

Qy    305 KDPKPVSGQKCEQICNEPCEEVVLKRSSKSKRKTDKKLMKKQQHSKKRTAQADVSDAKLC 364

Db   2231 ETIREDADQPMKPSQAESRRESIAESIKASSPRDEKSPLASKEASRPGSVAESI---KYD 2287

Qy    365 RRKPKKVRLLSEIINANQVEDSRSDEVHRENAADPCEDDRSTIPVPMEVSMDIPV-SNHT 423

Db   2288 LDKPQIIK-----------DDKSTEHSRRESL-----EDKSAVTSEKSVSRPLSVASDHE 2331

Qy    424 VG---EDGLKSS---KNKTKRKY-----------------SDVVDDGS---SLMNWLNG 456

Db   2332 AAVAIEDDAKSSISPKDKSRPGFVAETVSSPIEEATMEFSKIEVVEKSSLALSLQGGSGG 2391

Qy    457 KKK--------RTGSVHHTVAHPAGNLSNKKVTPTAS-----TQHDDENDTENGLDTNMH 503

Db   2392 KLQTDSSPVDVAEGDFSHAVA------SVSTVTPTLTKPAELAQIGAAKTVSSPLDEALR 2445

Qy    504 KTDVCQHVS------EISTQRCSSKGKTAGLSKGKTH---------SAASTKYGGESTRN 548

Db   2446 TPSAPEHISRADSPAECASEEIASQDKSPQVLKESSRPAWVAESKDDAAQLKSSVEDLRS 2505

Qy    549 GQNIHVLSAE---DQCQMETENSVLSHS----AKVSPAEH-----DIQIMSDLHEQSLP- 595

Db   2506 ----PVASTEISRPASAGETASSPIEEAPKDFAEFEQAEKAVLPLTIELKGNLPTLSSPV 2561

Qy    596 -------KKKKKQKLEVTREKQTMIDDIPMDIVELLAKNQHERQLMTETDCS-------D 641

Db   2562 DVAHASVQPAELSKVDIEKTASSPIDEAPKSLIGSPAEERPESPAESAKDAAESVEKSKD 2621

Qy    642 INR---IQSKTTAD-----------------DD------------------------- 654

Db   2622 ASRPPSVVESTKADSTKGDISPSPESVLEGPKDDVEKSKESSRPPSVSASITGDSTKDVS 2681

Qy    655 ---CVIVAAKDGSDYA----------SSVFDT--NSQQKSLASQSTQKELQGHLALTTQE 699

Db   2682 RPASVVESVKDEHDKAESRRESIAKVESVIDEAGKSDSKS-SSQDSQKDEKSTLA-SKEA 2739

Qy    700 SPHPQNFQSTQEQQTHLRMEEMVTIAASSPLFSHHDDQYIAEAPTEHWGRKDAKKLTWEQ 759

Db   2740 SRRESVVESSKDDAEKSESRPESVIASGEPV-----------PRESKSPLDSKDTS--- 2784

Qy    760 FKATTRNSPAATCGAQFRPGIQAVDLTSTHVMGSSSNYASRQPVIAPLDRYAERAVNQVH 819

Db   2785 ----------------RPGSVVESVTAE---DEKSEQQSRRESVAESVKADTKKDGKSQ 2824
```

```
Qy     820 ARNFPSTIATMEASKLCDRRNAGQVVLYPKESMPATH--LLRMMDPSTLASFPNYGTSSR 877
           : ||::   |  |  | :   :     ::|:  :|  :  | | | :          ||
Db    2825 EASRPSSVD--ELLKDDDEKQESR-----RQSITGSHKAMSTMGDESPMDKADKSKEPSR 2877

Qy     878 NQ--MESQLHNSQYAHNQYKGSTSTSYGSNLNG------KIPLTFEDLSRHQLHDLHRPL 929
            :   ||  | :         ||   |   ::         | ||  :::|| :
Db    2878 PESVAESIKHENTKDEESPLGSRRDSVAESIKSDITKGEKSPLPSKEVSRPE-------- 2929

Qy     930 RPHPRVGVLGSLLQKEIANWSENCGTQSGYKLGVSTGITSHQMNRKEHFEALNSGMFSAK 989
                 |:||:  ::  :  |:       |   |   ||   : |||    :
Db    2930 ------SVVGSIKDEKAESRRES--VAESVKPESSKDATSAPPS-KEHSRPES------- 2973

Qy     990 WNALQLGSVSSSADFLSARN-SIAQSWTRGKGKMV 1023
               |||:    |  ::|  |:| |    |  :|
Db    2974 ----VLGSLKDEGDKTTSRRVSVADSIKDEKSLLV 3004


RESULT  13
T34418
hypothetical protein F12F3.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T34418
R;Fulton, B.; Wohldmann, P.
submitted to the EMBL Data Library, July 1998
A;Description: The sequence of C. elegans cosmid F12F3.
A;Reference number: Z21521
A;Accession: T34418
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3488 <FUL>
A;Cross-references: EMBL:U80022; PIDN:AAC25885.1; GSPDB:GN00023; CESP:F12F3.3
A;Experimental source: strain Bristol N2; clone F12F3
C;Genetics:
A;Gene: CESP:F12F3.3
A;Map position: 5
A;Introns: 281/3; 332/1; 562/3; 600/3; 1866/3; 1944/3; 3393/1


  Query Match           3.7%;  Score 204;  DB 2;  Length 3488;
  Best Local Similarity   18.5%;  Pred. No. 0.013;
  Matches  214;  Conservative  161;  Mismatches  419;  Indels  364;  Gaps   46;

Qy      31 VLELTATPRQDAAAEAGVDEPAQHQCEHFSIRGYVALLQ---KKDP---KFCSLSRIFHD 84
           :::   ||  :    |  || | :    ::  ||  :   || |   |  ||:   |
Db     368 IIKKKETPESEQVTAA---EPEQQKISEVDVQS-VAETEVGAKKKPDAEKPTDLSKAKKD 423

Qy      85 --QKKCDEHKASSSPFSVAKFRRWDCSKCLDKLKTSDNGTAPRTLPAKQNGT-------- 134
             || || :||:   |  :    |        |||     :|:  |:  |
Db     424 SKSKKSDEPEASTEEKSTTEKPTND--------KTSKKSAEKKTVKPKKEVTGKPLEAKK 475

Qy     135 -------------------SDGCSITFV-RSTFVPASVGSQKVSPSTQSSQ-------- 165
                              :||     : :: |:|  : |:   |||  |:
Db     476 PVEDKKDASQPSSSKESSPPTDGKKKKQIPKALFIPDEIGSRFGDPSTMHSETNITTTIR 535

Qy     166 GK--NADRST-----LPKSV---------------------QEGNDSKCNAPSG----- 191
           |:  :||  |      |  ||                   |  | |    |
Db     536 GREGSADAKTPLVEPLSASVSMKVFTLVESAKEKAEFSFKRRSETPDDKSRKKEGLPPAK 595

Qy     192 ----------KNGAAEANTDSPMKDL-------QGPA-QNYDVAANVSE----DNTSVDV 229
                    :  : ||  :|  |::       | |: :|      | |      |  ||
Db     596 KSEKKDEVTAEKQSTEALIESKKKEVDESKISEQQPSDKNKSEVVGVPEKAAGPETKKDV 655

Qy     230 GALPEVP-QITWHIEVNGADQPPSTPKLSEVVLKRNEDENGKT------------EETLV 276
             : ||| : |   :   :|   |:      |||  :|:  |:             ::| |
Db     656 SEIEEVPKKKTIKKKTEKSD---SSISQKSNVLKPADDDKSKSDDVTDKSKKTTEDQTKV 712

Qy     277 AEQCNLTKDPNPMSGKERDQVAEQCNLTKDPKPVSGQKCEQICN--------------E 321
           |   | |  :     | : | :     |  | |  :| |:  :                :
Db     713 ATDSKLEKAADTTKQIETETVVDD----KSKKKVLKKKTEKSDSFISQKSETPPVVEPTK 768

Qy     322 PCEEVVLK----RSSKSKRKTDKKLMKKQQHSKKRTAQ-------------ADVSDAKL 363
           | |     |      :| ::: |  | :: : : |: |               |:|  ||
Db     769 PAESEAQKIAEVNKAKKQKEVDDNLKREAEVAAKKIADEKLKIEAEANIKKTAEVEAAKK 828

Qy     364 CRRKPKKVRLLSEIINANQVED------------SRSDEVHRE---NAADPCEDDRSTI 407
           : | ::::| :|:::     :             : :| | ::   |  :  |  : :
Db     829 QKEKDEQLKLETEVVSKKSAAEKLELEKQAQIKKAAEADAVKKQKELNEKNKLEAAKKSA 888

Qy     408 PVPMEVSMDIPVSNHTVGEDGLKSSKNK--------------TKRKYSDVVDDGSSLMN 452
              :::  :    :  | |: :|  : |               | :|  |  | |::
Db     889 ADKLKLEEESAAKSKKVSEESVKFGEEKKTKAGEKTVQVESEPTSKKTIDTKDVGATEPA 948

Qy     453 WLNGKK--------KRTGSVHHTVAHPAGNLSNKK--------------VTPTASTQHD 489
              ||         |   |:    |  :  :| :|              ||    ::
Db     949 DETPKKKIIKKKTEKSDSSISQKSATDSEKVSKQKEQDEPTKPAVSETQMVTEADKSKKQ 1008

Qy     490 DENDTENGLDTNM-----------HKTDVCQHVSEISTQRCSSKGKTAGLSKGKTHSAAS 538
           | | :  ||  :            | |  : : ::|    : | |     | |  | :
Db    1009 KETDEKLKLDAEIAAKTKQEADEKSKLDAQEKIKKVSEDDAARKEKELN-DKLKLESEIA 1067

Qy     539 TKYGG-------ESTRNGQNIHVLSA------EDQCQMETENSVLSHSAKVSPAEHDIQI 585
           ||          |  :   :    | :|      ::| :::|| :    :|:    |  ||
Db    1068 TKKASADKLKLEEQAQAKKAAEVEAAKKQKEKDEQLKLDTEAASKKAAAEKLELEKQAQI 1127

Qy     586 MSDLHEQSLPKKK---KKQKLEVTRE---------------KQTMIDDIPMDIVELLAK 626
                 :: |:|   :| |||  ::                |||: :   :|  :  ||
Db    1128 KKAAGADAVKKQKELDEKNKLEANKKSAAGKLKIEEESAAKSKQTVEEQAKLD-AQTKAK 1186

Qy     627 NQHERQLMTETDCSDINRIQSKTTADD------------------------DCVIVAAK 661
              |:|   | |        :|| | |:                        |: :|
Db    1187 TA-EKQTKLEKDEKSTKESESKETVDEKPKKKVLKKKTEKSDSSISQKSETSKTVVESAG 1245

Qy     662 DGSDYASSVFDTNSQQK---------------------------------SLASQSTQK 687
                 | |   :||                                   : :   ||
Db    1246 PSESETQKVADAARKQKETDEKQKLEAEITAKKSADEKSKLEAESKLKKAAEVEAAKKQK 1305

Qy     688 ELQGHLALTTQESPHPQNFQSTQ-EQQTHLRMEEMVTIAASSPLFSHHDDQYIAEAPTEH 746
           |    | | |: :      :  : |:|:|::    |  |          :  :|| |:
Db    1306 EKDEQLKLDTEAASKKAAAEKLELEKQSHIKKAAEVD-AVKKQKELEEKQRLESEAATK- 1363

Qy     747 WGRKDAKKLTWEQFKATTRNSPAATCG----------AQFRPGIQAVDLTSTHVMGSSS 795
             : ||:||  |: |       ||              || :  ::      |       |
Db    1364 --KADAEKLKLEEQK-----KKAAEIALIEIQKEQEKLAQEQSRLEDEAKKSAEKQKLES 1416

Qy     796 NYASRQPVIAPLDRYAERAVNQV-------------HARNFPSTI---ATMEAS-KLCD 837
              |:|   || :   |:   :|               :::  ||:    |:|:   | :
Db    1417 ETKSKQTEEAPKESVDEKPKKKVLKKKTEKSDSSISQKSKSAKSTVDAAETLESDFNLVE 1476

Qy     838 RRNAGQVVLYPKESMPAT 855
           ::   :|   | ||  ||
Db    1477 KKTVQKVEQSPDESTSAT 1494


RESULT  14
F90073
hypothetical protein SA2447 [imported] - Staphylococcus aureus (strain N315)
C;Species: Staphylococcus aureus
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C;Accession: F90073
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; O
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K
 C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A;Title: Whole genome sequencing of meticillin-resistant Stapylococcus aureus.
A;Reference number: A89758; MUID:21311952; PMID:11418146
A;Accession: F90073
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2271 <KUR>
A;Cross-references: GB:BA000018; PID:g13702612; PIDN:BAB43752.1; GSPDB:GN00149
A;Experimental source: strain N315
C;Genetics:
A;Gene: SA2447.
```

Query Match 3.7%; Score 202.5; DB 2; Length 2271;
Best Local Similarity 18.3%; Pred. No. 0.0088;
Matches 182; Conservative 138; Mismatches 415; Indels 259; Gaps 34;

```
Qy      90 EHKASSSPFSVAKFRRWDCSKCL-DKLKTSDNGTAPRTLPAKQNGTSDGCSITFVRSTFV 148
Db    1025 DSKSMSVSSSMSTSQSGSTSESLSDSQSTSDSDSKSLSLSTSQSGSTS-------TSTST 1077

Qy     149 PASV---GSQKVSPSTQSSQGKNADRSTLPKSVQEGNDSKCNAPSGKNGAAEANTDSPMK 205
Db    1078 SASVRTSESQSTSGSMSASQSDSMSIST------SFSDSTSDSKSASTASSESISQSAST 1131

Qy     206 DLQGPAQNYDVAANVSEDNTSVDVGALPEVPQITWHIEVNGADQPPSTPKLSEVVLKRNE 265
Db    1132 STSGSVSTSTSLSTSNSERTSTSV----------------SDSTSLSTSESDSISESTS 1174

Qy     266 DENGKTEETLVAEQCNLTKDPNPMSGKERDQVAEQCNLTKDPKPVSGQKCEQICNEPCEE 325
Db    1175 TSDSISEAISASESTSISLSESNSTSDSESQSASAFLSESLSESTSESTSESVSSSTSES 1234

Qy     326 VVLKRSSKSKRKTDKKLMKKQQHSKKRTAQADVSDAKLCRRK---------------- 367
Db    1235 TSLSDSTSESGSTSTSLSNSTSGSASISTSTSISESTSTFKSESVSTSLSMSTSTSLSNS 1294

Qy     368 ------------PKKVRLLSEIINANQVEDSRSDEVH---------RENAADPCEDDRS 405
Db    1295 TSLSTSLSDSTSDSKSDSLSTSMSTSDSISTSKSDSISTSTSLSGSTSESESDSTSSSES 1354

Qy     406 TIPVPMEVSMDIPVSNHTVGEDGLKSSKNKTKRKYSDVVDDGSSLMNWLNGKKKRTGSVH 465
Db    1355 KSD---STSMSISMSQSTSG-----STSTSTSTSLSDSTSTSLSLSASMN----QSGVDS 1402

Qy     466 HTVAHPAGNLSNKKVTPTASTQHDDENDTENGLDTNMHKTDVCQHVSEI-STQRCSSKGK 524
Db    1403 NSASQSASN------STSTSTSESDSQST---------STYTSQSTSQSESTSTSTSLSD 1447

Qy     525 TAGLSK-----GKTHSAASTKYGGESTRNGQNIHVLSAEDQCQ---------METENS-- 568
Db    1448 STSISKSTSQSGSTSTSASLS-GSESESDSQSISTSASESTSESASTSLSDSTSTSNSGS 1506

Qy     569 -----VLSHSAKVSPAEHDIQIMSDLHEQSLPKKKKKQKLEVTREKQTMIDDIPMDIVEL 623
Db    1507 ASTSTSLSNSASASESDSSSTSLSDSTSASMQSSESD-----SQSTSASLSD-------- 1553

Qy     624 LAKNQHERQLMTETDCSDINRIQSKTTADDDCVIVAAKDGSDYASSVFDTNSQQKSLA-S 682
Db    1554 --------SLSTSTS-NRMSTIASLSTS------VSTSESGSTSESTSESDSTSTSLSDS 1598

Qy     683 QSTQKELQGHLALTTQESPHPQNFQSTQEQQTHLRMEEMVTIAASSPLF--SHHDDQYIA 740
Db    1599 QSTSR--------STSASGSASTSTSTSDSRS--------TSASTSTSMRTSTSDSQSMS 1642

Qy     741 EAPTEHWGRKDAKKLTWEQFKATTRNSPAATCGAQFRPGIQAVDLTSTHVMGSSSNYASR 800
Db    1643 LSTSTSTSMSDSTSLS-DSVSDSTSDSTSASTSGSMSVSISLSDSTSTSTSASE------ 1695

Qy     801 QPVIAPLDRYAERAVNQVHARNFPSTIATMEASKLCDRRNAGQVVLYPKESMPATHLLRM 860
Db    1696 ---------------------------VMSAS-ISDSQSMSESV--------------- 1711

Qy     861 MDPSTLASFPNYGTSSRNQMESQLHNSQYAHNQYKGSTSTSYGSNLNGKIPLTFEDLSRH 920
Db    1712 ------------NDSESVSES---NSESDSKSMSGSTSVSDSGSLS------------- 1742

Qy     921 QLHDLHRPLRPHPRVGVLGSLLQKEIANWSENCGTQSGYKLGVSTGITSHQMNRKEHFEA 980
Db    1743 ----VSTSLRKSESVSESSSLSGSQ--SMSDSVSTSDSSSLSVSTSLRSSE-SVSESDSL 1795

Qy     981 LNSGMFSAKWNALQLGSVSSSADFLSARNSIAQS 1014
Db    1796 SDSKSTSGSTSTSTSGSLSTSTS-LSGSESVSES 1828
```

RESULT 15
B71609
hypothetical protein PFB0680w - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C;Accession: B71609
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.
.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H
Science 282, 1126-1132, 1998
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A;Reference number: A71600; MUID:99021743
A;Accession: B71609
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-665 <GAR>
A;Cross-references: GB:AE001410; GB:AE001362; NID:g3845245; PIDN:AAC71925.1; PID:g384
A;Experimental source: clone 3D7
C;Genetics:
A;Gene: PFB0680w

Query Match 3.6%; Score 201.5; DB 2; Length 665;
Best Local Similarity 17.8%; Pred. No. 0.0019;
Matches 135; Conservative 126; Mismatches 299; Indels 197; Gaps 28;

```
Qy      64 YVALLQK---KDPKFCSLSRIFHD-----QKKCDEHKASSSPFSVAKFRRWDCSKCLDKL 115
Db       8 FLAVLAKGFLRNKEHANLINSYNDIVEDINIKKEEKSSSEPPFIPIKNK-------IDNV 60

Qy     116 KTSDN----------------GTAPRTLPAKQNGTSDGCSITFVRSTFVPASVGSQKVS 158
Db      61 HTKNNNQYNLHNNKSNKTHLTYGTHTSFL---QNCTINDC---------VDVDNKDSEIN 108

Qy     159 PSTQSSQGKNADRSTLPKSVQEGN---------DSKCNAPSGKNGAAEANTDSPMKDLQG 209
Db     109 NITKEKDDNNNNNGT--KQIEEKNKINKSDLHRQNELNLQSGKNEQDINKNEKGKQDISN 166

Qy     210 P-AQN-YDVAANVS----------------EDNTSVDVGALPEVPQITWH-IEVNGADQ 249
Db     167 SNAENKKDVKEGVKELEEKKKEEKISDDHKVEENKKSDDHKVEENKKSDDHKVEENKKSD 226

Qy     250 PPSTPKLSEVVLKRNEDENGKTEETLVAEQCNLTKDPNPMSGKE-------RDQVAEQCN 302
Db     227 DHKIEEVKKVEEHEEDEEEDKKEKKSENKNKDENKDENDEDNDEISDEDEVDDDVEEDKN 286

Qy     303 ----LTKDPKPVSGQKCEQICNEPCEEVVLKRSSKSKRK---------TDKKLMKKQQHS 349
Db     287 ENDDIDDDKKETDKTHLEEEENEIIEKEFSDKKKNGKNKDTKKEKSKDTEKEKSKDIEKE 346

Qy     350 KKRTAQADVSDAKLCRRKPKKVRLLSEIINANQVEDSRSDEVHRENAADPCEDDRSTIPV 409
Db     347 KSKDKEKEKSKDKEKEKGKDKEKEKSKDI---EKEKEKDKDIEKEKSKDTAKEKEKDKDI 403

Qy     410 PMEVSMDIPVSNHTVGEDGLKSSKNKTKRKYSDVVDDGSSLMNWLNGKKKRTGSVHHTVA 469
Db     404 EKEKSKDM---------EKLKNKQNDEKKK-----DD--------NEKKKNDKQDIH--- 438

Qy     470 HPAGNLSNKKVTPTASTQHDDENDTENGLDTNMHKTDVCQHVSEISTQRCSSKGKTAGLS 529
Db     439 -------------------DDNDDENDMEEIEENDDEEDEDEDMENKKKKKKGKNGNEN 478

Qy     530 KGKTHSAASTKYGGESTRNGQNIHVLSAEDQCQMETENSVLSHSAKVSPAE---HDIQIM 586
Db     479 GNENGSENGNENGNENGNENENKNESENENENENENGNENENEKENEKDKNIKEIENV 538

Qy     587 SDLHEQSLPKKKKKQKLEVTREKQTMIDDIPMDIVELLAKNQHERQLMTETDCSDINRIQ 646
Db     539 TNANKENYEKINKNSEITITKSN----------IDIYNNNRN----------NDIDKVN 577

Qy     647 SKTTADDDCVIVAAKDGSDYASSVFDTNSQQKSLASQSTQKELQGHLALTTQESPH---P 703
Db     578 --------------------NHIF--TNQQKKHNLHNEQNKFNETLNVSTNHKNHYEEK 614
```

```
Qy     704 QNFQSTQ---EQQTHLRMEEMVTIAAS-SPLFSHHDD 735
           : ::|     ::: |  :  | ||  : :   ||| |
Db     615 KKYESNMFNVDKRMHKNLTSMDTILHNLNDKLSHHKD 651
```

Search completed: September 16, 2002, 22:25:09
Job time: 6498 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: September 16, 2002, 22:24:06 ; Search time 37.97 Seconds
(without alignments)
1077.866 Million cell updates/sec

Title: US-09-828-068-2
Perfect score: 5526
Sequence: 1 MEIVAVDQEGARVVGTNCML.....NKNPADFTTISNDNEYMDYR 1057

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	225	4.1	1453	1	NKCR_MOUSE	P30415 mus musculu
2	217	3.9	1165	1	YNF4_YEAST	P53950 saccharomyc
3	216.5	3.9	2476	1	ATRX_MOUSE	Q61687 mus musculu
4	208.5	3.8	2492	1	ATRX_HUMAN	P46100 homo sapien
5	200.5	3.6	1849	1	IGA4_HAEIN	P45386 haemophilus
6	197	3.6	708	1	ICAL_HUMAN	P20810 homo sapien
7	196.5	3.6	2468	1	MAPB_HUMAN	P46821 homo sapien
8	194.5	3.5	3924	1	ANK2_HUMAN	Q01484 homo sapien
9	191	3.5	1233	1	YFI6_YEAST	P43597 saccharomyc
10	190.5	3.4	2156	1	RP1_HUMAN	P56715 homo sapien
11	184	3.3	2459	1	MAPB_RAT	P15205 rattus norv
12	183	3.3	875	1	ZIP1_YEAST	P31111 saccharomyc
13	182.5	3.3	2464	1	MAPB_MOUSE	P14873 mus musculu
14	181.5	3.3	1972	1	P531_HUMAN	Q12888 homo sapien
15	181.5	3.3	2663	1	CENE_HUMAN	Q02224 homo sapien
16	181	3.3	2295	1	WDR9_HUMAN	Q9nsi6 homo sapien
17	180.5	3.3	1727	1	ALM1_SCHPO	Q9utk5 schizosacch
18	179	3.2	1863	1	BRC1_HUMAN	P38398 homo sapien
19	178.5	3.2	730	1	GLN3_YEAST	P18494 saccharomyc
20	177.5	3.2	1462	1	NKCR_HUMAN	P30414 homo sapien
21	175.5	3.2	816	1	YG3A_YEAST	P53278 saccharomyc
22	175.5	3.2	1878	1	BRC1_CANFA	Q95153 canis famil
23	175	3.2	633	1	MLH_TETTH	P40631 tetrahymena
24	174.5	3.2	1132	1	YKK5_YEAST	P34250 saccharomyc
25	174	3.1	2175	1	HMCU_DROME	P10180 drosophila
26	172.5	3.1	1447	1	BUD4_YEAST	P47136 saccharomyc
27	172	3.1	1359	1	ATRX_CAEEL	Q9u7e0 caenorhabdi
28	171.5	3.1	731	1	BAF1_YEAST	P14164 saccharomyc
29	171	3.1	1395	1	SP41_YEAST	P38904 saccharomyc
30	171	3.1	2245	1	MYSJ_DICDI	P54697 dictyosteli
31	170.5	3.1	1164	1	KEL1_YEAST	P38853 saccharomyc
32	170.5	3.1	1189	1	YJH6_YEAST	P47035 saccharomyc
33	170.5	3.1	1257	1	RBB1_HUMAN	P29374 homo sapien
34	169	3.1	1781	1	AKAC_HUMAN	Q02952 homo sapien
35	169	3.1	1969	1	MYSA_CAEEL	P12844 caenorhabdi
36	168.5	3.0	1805	1	NEST_RAT	P21263 rattus norv
37	168.5	3.0	2194	1	SC16_YEAST	P48415 saccharomyc
38	167.5	3.0	997	1	BIR1_SCHPO	O14064 schizosacch
39	167.5	3.0	1403	1	YDF3_SCHPO	Q10475 schizosacch
40	167.5	3.0	2845	1	APC_MOUSE	Q61315 mus musculu
41	166.5	3.0	513	1	DMP1_HUMAN	Q13316 homo sapien
42	166	3.0	2314	1	AKA6_RAT	Q9wvc7 rattus norv
43	165.5	3.0	1014	1	HEX2_YEAST	Q00816 saccharomyc
44	165	3.0	918	1	YMJB_CAEEL	P34487 caenorhabdi
45	165	3.0	1391	1	MST2_DROHY	Q08696 drosophila

ALIGNMENTS

RESULT 1
NKCR_MOUSE
ID NKCR_MOUSE STANDARD; PRT; 1453 AA.
AC P30415;
DT 01-APR-1993 (Rel. 25, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE NK-tumor recognition protein (Natural-killer cells cyclophilin-
DE related protein) (NK-TR protein).
GN NKTR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93133824; PubMed=8421688;
RA Anderson S.K., Gallinger S., Roder J., Frey J., Young H.A.,
RA Ortaldo J.R.;
RT "A cyclophilin-related protein involved in the function of natural
RT killer cells.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:542-546(1993).
RN [2]
RP REVISIONS TO C-TERMINUS.
RC STRAIN=BALB/C; TISSUE=Blood;
RA Anderson S.K.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
CC -!- FUNCTION: COMPONENT OF A PUTATIVE TUMOR-RECOGNITION COMPLEX.
CC INVOLVED IN THE FUNCTION OF NK CELLS.
CC -!- SIMILARITY: CONTAINS 1 CYCLOPHILIN-LIKE PPIASE DOMAIN.
CC ---

CC ---
DR EMBL; L04289; AAA37500.2; ALT_INIT.
DR HSSP; Q27450; 1A33.
DR MGD; MGI:97346; Nktr.
DR InterPro; IPR002130; CSA_PPIase.
DR Pfam; PF00160; pro_isomerase; 1.
DR PRINTS; PR00153; CSAPPISMRASE.
DR PROSITE; PS00170; CSA_PPIASE_1; 1.
DR PROSITE; PS50072; CSA_PPIASE_2; 1.
KW Cyclosporin; Isomerase; Rotamase; Repeat; Transmembrane.
FT DOMAIN 1 176 PPIASE, CYCLOPHILIN-TYPE.
FT DOMAIN 222 241 ARG/LYS-RICH (BASIC).
FT DOMAIN 422 459 ARG/LYS-RICH (BASIC).
FT DOMAIN 964 1003 ARG/LYS-RICH (BASIC).
FT DOMAIN 198 273 ARG/SER-RICH.
FT DOMAIN 468 565 ARG/SER-RICH.
FT DOMAIN 658 812 ARG/SER-RICH.
FT DOMAIN 1303 1453 ARG-SER TANDEM REPEAT-RICH.

```
SQ   SEQUENCE   1453 AA;  163439 MW;  DF1173FF814B283E CRC64;

  Query Match             4.1%;  Score 225;  DB 1;  Length 1453;
  Best Local Similarity   19.5%;  Pred. No. 0.00028;
  Matches  239;  Conservative  169;  Mismatches  459;  Indels  356;  Gaps   52;

Qy      93 ASSSPFSVAKFRRWDCSKCLDKLKTSDNGTAPRTLPAKQNGTSDGCSITFVRSTFVPASV 152
Db     160 AASRPY--ADVRVIDCGVLATKL-TKDVFEKKRKKPTCSEG----------------SD 199

Qy     153 GSQKVSPSTQSSQGKNADRSTL--------PK----------SVQEGNDSKCNAPSGK 192
Db     200 SSSRSSSSSESSSESEVERETIRRRRHKRRPKVRHAKKRFKEMSSSEEPRRKRTVSPEGY 259

Qy     193 NGAAEANTDSPMKDLQGPAQNYDVAANVSEDNTSVDVGAIPEVPQ----------ITWH 241
Db     260 SERSDVN----------EKRSVDSNTKREKPVVRPEEIPPVPENRFLLRRDMPAITVE 307

Qy     242 IEVN-------GADQPPS----------------TPKLSEVVLKRNEDENGKTEETLVA 277
Db     308 PEQNIPDVAPVVSDQKPSVSKSGRKIKGRGTIRYHTPPRSRSHSESKDDDSSETPPHWKE 367

Qy     278 EQCNLT--KDPNPMSGKERDQVAEQCNLTKDPKPVSGQKCEQICNEPCEEVVLKRSSKSK 335
Db     368 EMQRLRAYRPPSGEKWSKGDKLSDPCSSRWDERSLSQRSFSWSYNGYYSDLSTARHSDGH 427

Qy     336 RKTDKKLMKKQQHSKKRTAQADVSDAKLCR--RKPKKVRIL----------SEIINANQV 383
Db     428 HKKHRK-EKKFKHKKKAKKQ------KHCRRHRQTKKRRIVMPDLEPSRSPTHRMKSSCV 480

Qy     384 EDSR----SDEVHRENAADPCEDDRSTIPVPMEVSMDIPVSNHTVGEDGLKSSK------ 433
Db     481 RERRSRASSSSSHHSSKRDWSKSDQDDGSASTHSSRDSYFSKSHSRSDSRGSSRSRAVSK 540

Qy     434 ------NKTKRKYS----------DVVDDGSSLM----- 451
Db     541 SSSRSLNRSKSRSSSRSGPRRTSISPKKPAQLSENKPVKTEPLRPSVPQNGNVLVQPVAA 600

Qy     452 ---------------NWLNGKK----------KRTGSVHHTVAHPAGNLSNKKVTPTAS 485
Db     601 ENIPVIPLSDSPPPSRWKPGQKPWKPSYERIQEMKAKTTHLLPVQSTYSLTNIKATVSSS 660

Qy     486 TQHDDENDTENGLDTNMHKTDVC-----QHVSEISTQRCSSKGKTAGLSK---------- 530
Db     661 SYHKREKPSESDGSAYSKYSDRSSGSSGRSGSKSSRSRSSSRSYTRSRSRSLPTSRSLSR 720

Qy     531 ---GKTHSAASTKYGGESTRNGQNIHVLSAEDQCQMETEN---SVLSHS---------- 573
Db     721 SPSSRSHSPNKYSDGSQHSRSSSYTSVSSDDGRRAMFRSHRKKSVTSHKRHRSNSEKTLH 780

Qy     574 ------------AKVSPAEHDIQIMSD---LHEQ--SLPKKKKKQKLEVTREKQTMIDD 615
Db     781 SKYVRGREKSSRHRKYSESRSSLDYTSDSDQSHVQVYSAPEKEKQGKVEALNDKQ----- 835

Qy     616 IPMDIVELLAKNQHERQLMTETDCSDINRIQSKTTADDDCVIVAAKDGSDYASSVFDTNS 675
Db     836 ---------GKGREEGKPKPEWECPRSKKENSEDHSRDDSV----SKGKNCAGSKWDSES 882

Qy     676 QQKSLASQSTQKELQGHLALTTQESPHPQNFQSTQEQQTHLRMEEMVTIAASSPLFSHHD 735
Db     883 NSEQDVTKSRKSDPRRG---SEKEEGEASSDSESEVGQSHIK--------AKPP------ 925

Qy     736 DQYIAEAPTEHW--GRKDAKKLTWEQFKATTRNSPAATCGAQFRPGIQAVDLTSTHVMGS 793
Db     926 ----AKPPTSTFLPGSDGAWKSRRPQSSASESESSCSNLG-NIRGEPQKQKHSKDDLKGD 980

Qy     794 SSNYASRQPVIAPLDRYAERAVNQVHARNFPSTIATMEASKLCDRRNAGQVVLYPKESMP 853
Db     981 HTKRA-REKSKAKKDK-KHKAPKRKQAFHWQPPLEFGDDIE--EEMNGKQVTQDPKEKRH 1036

Qy     854 AT----------HLLRMMDPSTLASFPNYGTSSRNQMESQLH--NSQYAHNQYKGSTST 900
```

```
Db    1037 VSEKCEAVKDGIPNVEKTCDEGSSPSKPKKGTLEQDPLAEGGHDPSSCPAPLKVEDNTAS 1096

Qy     901 S---------YGSNLNGKIPLTFED--------------------LSRHQLHDLHRPLRP 931
Db    1097 SPPSAQHLEEHGPG-GGEDVLQTDDNMEICTPDRTSPAKGEVVSPLANHRLDSPEVNIIP 1155

Qy     932 -------HPRVGVLGSLLQKEIANWSENCGTQSGYKLGVSTGITS--HQMNRKEHFEALN 982
Db    1156 EQDECMAHPRAG--GE--QESSMSESKTLG-ESGVKQDSSTSVTSPVETSGKKEGAEKSQ 1210

Qy     983 SGMFSAKWNALQ-LGSVS------SSADFLSARNSIAQSWTRG-----KGK-MVHPLDRF 1029
Db    1211 MNL-TDKWKPLQGVGNLSVSTATTSSALDVKALSTVPEVKPQGLRIEIKSKNKVRPGSLF 1269

Qy    1030 --VRQDICITNKNPADFTTISND 1050
Db    1270 DEVRKTARL-NRRPRNQESSSDD 1291


RESULT   2
YNF4_YEAST
ID   YNF4_YEAST     STANDARD;      PRT;  1165 AA.
AC   P53950;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   Hypothetical 128.1 kDa protein in OMP2-MSG5 intergenic region.
GN   YNL054W OR N2467 OR YNL2467W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / FY1676;
RX   MEDLINE=96021608; PubMed=8533472;
RA   Bergez P., Doignon F., Crouzet M.;
RT   "The sequence of a 44 420 bp fragment located on the left arm of
RT   chromosome XIV from Saccharomyces cerevisiae.";
RL   Yeast 11:967-974(1995).
RN   [2]
RP   ERRATUM.
RX   MEDLINE=97060022; PubMed=8904343;
RA   Bergez P., Doignon F., Crouzet M.;
RL   Yeast 12:297-297(1996).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U12141; AAA99658.1; -.
DR   EMBL; Z71330; CAA95925.1; -.
DR   SGD; S0004999; YNL054W.
KW   Hypothetical protein.
SQ   SEQUENCE   1165 AA;  128140 MW;  0414978AA3066CE7 CRC64;

  Query Match             3.9%;  Score 217;  DB 1;  Length 1165;
  Best Local Similarity   18.7%;  Pred. No. 0.00055;
  Matches  204;  Conservative  156;  Mismatches  416;  Indels  314;  Gaps   46;

Qy     155 QKVSPSTQSSQGKNADRSTLPKSVQEGNDSKCNAPSGKNGAAEANTDSPM--KDLQGPAQ 212
Db       6 RKLTVETETVEAPVANNLLL------SNNSNVVAPNPSIPSASTST-SPLHREIVDDSVA 58

Qy     213 NYDVAANVSE------DNTSVDVGALPEVPQITWHIEVNGADQPPSTPKLSEVVLKRNED 266
Db      59 TANTTSNVVQHNLPTIDNNLMDSDATSH-NQDHWHSDINRAGTSMST---SDIPTDLHLE 114
```

```
Qy     267 ENGKTEETLVAEQCNLTKDPNPMSGKERDQVAEQCNLTKDPKPVSGQKCEQICNEPCEEV 326
Db     115 HIGSVSST-NNNSNNALINHNPLSSH---------------------------------- 139

Qy     327 VLKRSSKSKRKTDKKLMKKQQHSKKRTAQADVSDAKLCRRKPKKVRLLSEIINANQVEDS 386
Db     140 -LSNPSSSLRNKKSSLLVASN-------PAFASDVELSKKKPAVISNNMPTSNIALYQTA 191

Qy     387 RSDEVHRENAADPCEDDR------------------STIPVPMEVSMDIPVSNHTVGED 427
Db     192 RSANIHGPSSTSASKAFRKASAFSNNTAPSTSNNIGSNTPPAPL-----LPLPS------ 240

Qy     428 GLKSSKNKTK---RKYSDVVDDGSSLM--NWLNGKKKRTGSVHHTVAHPAGNLSNKKV-T 481
Db     241 --LSQQNKPKIIERPTMHVTNSREILLGENLLDDTKAKNAPANST-THDNGPVANDGLRI 297

Qy     482 PTASTQHDDENDTENGLDTNMH-----KTDVCQHVSEISTQRCSS----------KGKTA 526
Db     298 PNHSNADDNENNNKMKKNKNINSGKNERNDDTSKICTTSTKTAPSTAPLGSTDNTQALTA 357

Qy     527 GLSKGKTHSAASTKYGGESTRNGQNIHVLSAEDQCQMETENSVLSHSAKVSPAEHDIQIM 586
Db     358 SVSSSNADNHNNNKKKTSSNNNGNNSNSASNKTNADIKNSNADLSASTSNNNAIND---- 413

Qy     587 SDLHEQSLPKKKKKQ-----------KLEVTREKQTMIDDIPMDIVELL-------AKNQ 628
Db     414 -DSHESNSEKPTKADFFAARLATAVGENEISDSEETFVYESAANSTKNLIFPDSSSQQQQ 472

Qy     629 HERQLMTETDCSDINRIQSKTTA---DDDCVIVAAKDGSDYAS---------SVFDTNSQ 676
Db     473 QQQQPPKQQQQQQNHGITSKISAPLLNNNKKLLSRLKNSRHISTGAILNNTIATISTNPN 532

Qy     677 QKSLASQSTQKELQGH-----LALTTQESPHPQNFQSTQEQQTHLRMEEMVTIAASSPLF 731
Db     533 LNSNVMQNNNNLMSGHNHLDELSSIKQEPPH-----QLQQQQPPMDVQSVDSYTSDNP-- 585

Qy     732 SHHDDQYIAEAPTEHWGRKDAKKLTWEQFKATTRNSPAATC------GAQFRPGIQAVDL 785
Db     586 ---DSNVIAKSPDKRSSLVSLSKVSPHLLSSTSSNGNTISCPNVATNSQELEPN---NDI 639

Qy     786 TSTHVMGSS----------SNYASRQ--------------PVIAPLDRYA--------ER 813
Db     640 STKKSLSNSTLRHSSANRNSNYGDNKRPLRTTVSKIFDSNPNGAPLRRYSGVPDHVNLED 699

Qy     814 AVNQVHARNFPSTIATMEASKLCDRRNAGQVVLYPKESMPATHLLRMMDPSTLASFPNYG 873
Db     700 YIEQPH--NYPTMQNSVKKDEFYNSRN---------NKFP--HGLNFYGDNNVIEEENNG 746

Qy     874 TSS---------------------------------RNQMESQLHNSQYAH-------- 891
Db     747 DSSNVNRPQHTNLQHEFIPEDNESDENDIHSMFYYNHKNDLETKPLISDYGEDEDVDDYD 806

Qy     892 ------NQYKGSTSTSYGSNLNGKIPL----TFEDL-------SRHQLHDLHRPLRPHPR 934
Db     807 RPNATFNSYYGSASNTHELPLHGRMPSRSNNDYYDFMVGNNTGNNNQLNE-YTPLRMKRG 865

Qy     935 VGVLGSLLQKEIANWSENCG-----TQSGYKLGVSTGITSHQM-NRKEHFEALNSGMFSA 988
Db     866 QRHL-SRTNNSIMNGSIHMNGNDDVTHSNINNNDIVGYSPHNFYSRKSPFVKVK------ 918

Qy     989 KWNALQLGSVSSS-------ADFLSARNSIAQSWTRGKGKMVHPLDRFVRQ------DIC 1035
Db     919 --NFLYLAFVISSLLMTGFILGFLLATNKELQDVD------VVVMDNVISSSDELIFDIT 970

Qy    1036 ITNKNPADFT 1045
Db     971 VSAFNPGFFS 980
```

```
RESULT   3
ATRX_MOUSE
ID   ATRX_MOUSE     STANDARD;      PRT;  2476 AA.
AC   Q61687;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Transcriptional regulator ATRX (X-linked nuclear protein)
DE   (Heterochromatin protein 2) (HP1 alpha-interacting protein) (HP1-BP38
DE   protein).
GN   ATRX OR XNP OR HP1BP2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98213653; PubMed=9545503;
RA   Picketts D.J., Tastan A.O., Higgs D.R., Gibbons R.J.;
RT   "Comparison of the human and murine ATRX gene identifies highly
RT   conserved, functionally important domains.";
RL   Mamm. Genome 9:400-403(1998).
RN   [2]
RP   SEQUENCE OF 325-1176 FROM N.A.
RX   MEDLINE=97133299; PubMed=8978696;
RA   le Douarin B., Nielsen A.L., Garnier J.-M., Ichinose H.,
RA   Jeanmougin F., Losson R., Chambon P.;
RT   "A possible involvement of TIF1 alpha and TIF1 beta in the epigenetic
RT   control of transcription by nuclear receptors.";
RL   EMBO J. 15:6701-6715(1996).
RN   [3]
RP   SUBCELLULAR LOCATION, AND ASSOCIATION WITH PERICENTROMERIC
RP   HETEROCHROMATIN.
RX   MEDLINE=20040663; PubMed=10570185;
RA   McDowell T.L., Gibbons R.J., Sutherland H., O'Rourke D.M.,
RA   Bickmore W.A., Pombo A., Turley H., Gatter K., Picketts D.J.,
RA   Buckle V.J., Chapman L., Rhodes D., Higgs D.R.;
RT   "Localization of a putative transcriptional regulator (ATRX) at
RT   pericentromeric heterochromatin and the short arms of acrocentric
RT   chromosomes.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:13983-13988(1999).
CC   -!- FUNCTION: COULD BE A GLOBAL TRANSCRIPTIONAL REGULATOR. MODIFIES
CC       GENE EXPRESSION BY AFFECTING CHROMATIN.
CC   -!- SUBUNIT: PROBABLY BINDS EZH2. BINDS ANNEXIN V IN A CALCIUM AND
CC       PHOSPHATIDYLCHOLINE/PHOSPHATIDYLSERINE-DEPENDENT MANNER (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR. ASSOCIATED WITH PERICENTROMERIC
CC       HETEROCHROMATIN DURING INTERPHASE AND MITOSIS, PROBABLY BY
CC       INTERACTING WITH HP1.
CC   -!- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 PHD-TYPE ZINC FINGER.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF026032; AAC08741.1; -.
DR   EMBL; X99643; CAA67962.1; -.
DR   MGD; MGI:103067; Xnp.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR000330; SNF2_N.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00176; SNF2_N; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
DR   SMART; SM00184; RING; 1.
KW   DNA repair; Nuclear protein; DNA-binding; Helicase; ATP-binding;
KW   Zinc-finger.
FT   ZN_FING     219    267       PHD-TYPE.
FT   NP_BIND    1579   1586       ATP (POTENTIAL).
FT   SITE       1704   1707       DEGH BOX.
```

```
FT   DOMAIN      319    322       POLY-SER.
FT   DOMAIN      735    738       POLY-SER.
FT   DOMAIN     1001   1004       POLY-GLU.
FT   DOMAIN     1130   1135       POLY-SER.
FT   DOMAIN     1182   1185       POLY-SER.
FT   DOMAIN     1238   1245       POLY-ASP.
FT   DOMAIN     1484   1487       POLY-GLU.
FT   DOMAIN     1924   1931       POLY-SER.
FT   DOMAIN     2205   2208       POLY-LYS.
FT   DOMAIN     2245   2248       POLY-GLU.
FT   DOMAIN     2403   2408       POLY-GLN.
SQ   SEQUENCE   2476 AA;  278601 MW;  90A42B790FC4FF4C CRC64;


  Query Match             3.9%;  Score 216.5;  DB 1;  Length 2476;
  Best Local Similarity   19.3%;  Pred. No. 0.0015;
  Matches  165;  Conservative  133;  Mismatches  346;  Indels  209;  Gaps   36;

Qy      83 HDQKKCDEHKASSSPFSVAKFRR-WDCSKCLDKLKTSDNGTAPRTLPAKQNGTSDGCSIT 141
           |:: |  :  :|:      : |:   ||     ::   :    | :  |   :|||
Db     844 HERAKTTQEGSSADDTGDTEGRQGGSCSIAGGSIEKVRSGVEFREMLCKPGVSSDGAEKP 903

Qy     142 FVRSTFVPASVGSQKVSP---------STQSSQGK-----NADRSTLPKSVQEGNDSKCN 187
            |:   |  |   ::||           | || :||       ||   ||  |    |:
Db     904 SVKEENV-NSPEDKRVSKTKEKTKHLRSRQSRKGKGGSSDGTDR--FPKKEQSDESSEGE 960

Qy     188 APSGKNGAAEANTDSPMKDLQGPAQNYDVAANVSEDNTSVDVGALPEVPQITWHIEVNGA 247
           :          :|  ||:|      :   : | | |       |||  :|    :  |
Db     961 KKQSRQRPGTKGKKAP--DLKGETLKREQEWDSSSDGTE----RLPEEEEIGPFSK--GI 1012

Qy     248 DQPPSTPKLSEVVLKRNEDENGKTEETL---------VAIQCNLTKDP---NPMSGKERD 295
            |  :    |   |: :|:: : :| |           : |: ::|    | :| :|:
Db    1013 KQSKTDTAGGEKKGKKWKDKSCEKKEELSDSVDKLPGKGDSCDSSEDKKTRNRVSLREKK 1072

Qy     296 QVA---------EQCNLTKDPKPVSGQKCEQICNEPCE-IVVLKRSSKSKRKT------- 338
           : :           :|: :   | : || |:      |   : : :|:| ||| |
Db    1073 RFSLPAKSPGKRPECSSSDTEKSLKGQCCDSTEKRPKRIDLRERRNSSSKRNTKEVKSAS 1132

Qy     339 ----------DKKLMKKQQHS-KKRTAQADVSDAKLCRREPKKVRLLSEIINANQVEDSR 387
                    | |  |||: | ||:|            |  ||: ::          :  |
Db    1133 SSSDAEGSSEDNKKQKKQRTSAKKKTGNTKEKKRNSLRATPKRKQV--------DITSSS 1184

Qy     388 SD--EVHRENAADPCEDDRSTIPVPMEVSMDIPVSNHT-----VGEDGLKSS-------- 432
           ||  :  : :| :   |::    |    |: :: : :||      |:: |  |
Db    1185 SDIGDDDQNSAGEESSDEQKIKP----VTENLVLPSHTGICQSSGDEALSKSVPATVDDD 1240

Qy     433 ------KNKTKRKY----------SDVVDDGSSLMNWLNGKKKRTGSVHHTVAHPAGNLS 476
                :|:  :|           ||  :|||       | ||| |           | |
Db    1241 DDDNDPENRIAKKMLLEEIKANLSSD--EDGSSDDEPDGGGKKRIGKQSEESPADDGELR 1298

Qy     477 NKKVTPTASTQHDDENDTENGLDTNMHKTDVCQHVSEISTQRCSSKGKTAGLSKGKTHSA 536
            :::          | :  |:      |:  : :|   :|      :  |    | | |
Db    1299 REQLAVNQVNSESDSDSEESKKPRYRHR--LLRHKLTLSIGESGEEKPT----KPKEHKE 1352

Qy     537 ASTKYGGESTRNGQNIHVLSAEDQCQMETENSVLSHSAKVSPAEHDIQIMSDLHEQSLPK 596
           |           |:|   :|:||   : : | :|                 :|:  ::  |:
Db    1353 A----------KGRNRRKVSSEDSEDTDFQESGVSEE-----------VSESEDEQRPR 1390

Qy     597 KKKKQKLEVTREKQTMIDDIPMDIVELLAKNQHERQLMTETDCSDINRIQS---KTTADD 653
            :  :| |:   :::              : :  |::  : | |  |:  |   |   |:
Db    1391 TRSAKKAELEENQRS-----------YKQKKKRRRIKVGEDSSSENKSHSEEDKKEGDE 1438

Qy     654 DCVIVAAKDGSDYASSVFDTNSQQKS----------LASQSTQKELQGHLALTTQESPH 702
           :       | |      | |   ||           |     : | |   ||  :|
Db    1439 E-------DEEDEDEDEEDENDDSKSPGKGRKKIRKILKIDKLRTETQN--ALKEEEERR 1489

Qy     703 PQNFQSTQEQQTHLRMEEMVTIAASS-------------------PLFSHHDDQYIAEA 742
            :  :  :|::   :: |:: |  :|                    ||   | :  |
Db    1490 KRIAERERERE---KLREVIEIEDASPTKCPITTKLVLDENEETKEPLVQVHRNMVIKLK 1546

Qy     743 PTEHWGRKDAKKLTWE---QFKATTRNSPAATCGAQFRPG----IQAVDLTSTHVMGSSS 795
           | :     |  :  |:   :    |: || : |      |    :| |    | ::
Db    1547 PHQ----VDGVQFMWDCCCESVEKTKKSPGSGCILAHCMGLGKTLQVVSFLHTVLLCDKL 1602

Qy     796 NYASRQPVIAPLD 808
           :: |   |: ||:
Db    1603 DF-STALVVCPLN 1614


RESULT   4
ATRX_HUMAN
ID   ATRX_HUMAN     STANDARD;      PRT;  2492 AA.
AC   P46100; P51068; Q15886; Q9NTS3; Q9H0Z1;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Transcriptional regulator ATRX (X-linked helicase II) (X-linked
DE   nuclear protein) (XNP) (Znf-HX).
GN   ATRX OR RAD54L OR XH2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 4 AND 5), AND VARIANT S-1860,
RP   AND VARIANTS ATR-X.
RX   MEDLINE=97123494; PubMed=8968741;
RA   Picketts D.J., Higgs D.R., Bachoo S., Blake D.J., Quarrell O.W.J.,
RA   Gibbons R.J.;
RT   "ATRX encodes a novel member of the SNF2 family of proteins: mutations
RT   point to a common mechanism underlying the ATR-X syndrome.";
RL   Hum. Mol. Genet. 5:1899-1907(1996).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORMS 2 AND 4).
RX   MEDLINE=97386582; PubMed=9244431;
RA   Villard L., Lossi A.-M., Cardoso C., Proud V., Chiaroni P.,
RA   Colleaux L., Schwartz C., Fontes M.;
RT   "Determination of the genomic structure of the XNP/ATRX gene encoding
RT   a potential zinc finger helicase.";
RL   Genomics 43:149-155(1997).
RN   [3]
RP   SEQUENCE OF 860-2492 FROM N.A.
RX   MEDLINE=95179111; PubMed=7874112;
RA   Stayton C.L., Dabovic B., Gulisano M., Gecz J., Broccoli V.,
RA   Giovanazzi S., Bossolasco M., Monaco L., Rastan S., Boncinelli E.,
RA   Bianchi M.E., Consalez G.G.;
RT   "Cloning and characterization of a new human Xq13 gene, encoding a
RT   putative helicase.";
RL   Hum. Mol. Genet. 3:1957-1964(1994).
RN   [4]
RP   PRELIMINARY PARTIAL SEQUENCE FROM N.A.
RX   MEDLINE=94214473; PubMed=8162050;
RA   Gecz J., Pollard H., Consalez G., Villard L., Stayton C.L.,
RA   Millasseau P., Khrestchatisky M., Fontes M.;
RT   "Cloning and expression of the murine homologue of a putative human
RT   X-linked nuclear protein gene closely linked to PGK1 in Xq13.3.";
RL   Hum. Mol. Genet. 3:39-44(1994).
RN   [5]
RP   SEQUENCE OF 2401-2492 FROM N.A., AND VARIANTS ATR-X.
RX   MEDLINE=95211835; PubMed=7697714;
RA   Gibbons R.J., Picketts D.J., Villard L., Higgs D.R.;
RT   "Mutations in a putative global transcriptional regulator cause X-
RT   linked mental retardation with alpha-thalassemia (ATR-X syndrome).";
RL   Cell 80:837-845(1995).
RN   [6]
RP   SEQUENCE OF 1375-2492 FROM N.A.
RA   Pearce A., Chapman J.;
RL   Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   EZH2 BINDING.
RX   MEDLINE=98167853; PubMed=9499421;
RA   Cardoso C., Timsit S., Villard L., Khrestchatisky M., Fontes M.,
RA   Colleaux L.;
RT   "Specific interaction between the XNP/ATR-X gene product and the SET
RT   domain of the human EZH2 protein.";
RL   Hum. Mol. Genet. 7:679-684(1998).
```

```
RN   [8]
RP   SUBCELLULAR LOCATION, AND ASSOCIATION WITH PERICENTROMERIC
RP   HETEROCHROMATIN.
RX   MEDLINE=20040663; PubMed=10570185;
RA   McDowell T.L., Gibbons R.J., Sutherland H., O'Rourke D.M.,
RA   Bickmore W.A., Pombo A., Turley H., Gatter K., Picketts D.J.,
RA   Buckle V.J., Chapman L., Rhodes D., Higgs D.R.;
RT   "Localization of a putative transcriptional regulator (ATRX) at
RT   pericentromeric heterochromatin and the short arms of acrocentric
RT   chromosomes.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:13983-13988(1999).
RN   [9]
RP   DISEASE.
RX   MEDLINE=20213147; PubMed=10751095;
RA   Villard L., Fontes M., Ades L.C., Gecz J.;
RT   "Identification of a mutation in the XNP/ATR-X gene in a family
RT   reported as Smith-Fineman-Myers syndrome.";
RL   Am. J. Med. Genet. 91:83-85(2000).
RN   [10]
RP   VARIANT ATR-X SER-1713.
RX   MEDLINE=97196774; PubMed=9043863;
RA   Villard L., Lacombe D., Fontes M.;
RT   "A point mutation in the XNP gene, associated with an ATR-X phenotype
RT   without alpha-thalassemia.";
RL   Eur. J. Hum. Genet. 4:316-320(1996).
RN   [11]
RP   VARIANT JM GLN-2131.
RX   MEDLINE=96224392; PubMed=8630485;
RA   Villard L., Gecz J., Mattei J.-F., Fontes M., Saugier-Veber P.,
RA   Munnich A., Lyonnet S.;
RT   "XNP mutation in a large family with Juberg-Marsidi syndrome.";
RL   Nat. Genet. 12:359-360(1996).
RN   [12]
RP   VARIANTS ATR-X.
RX   MEDLINE=97467722; PubMed=9326931;
RA   Gibbons R.J., Bachoo S., Picketts D.J., Aftimos S., Asenbauer B.,
RA   Bergoffen J., Berry S.A., Dahl N., Fryer A., Keppler K., Kurosawa K.,
RA   Levin M.L., Masuno M., Neri G., Pierpont M.E., Slaney S.F.,
RA   Higgs D.R.;
RT   "Mutations in transcriptional regulator ATRX establish the functional
RT   significance of a PHD-like domain.";
RL   Nat. Genet. 17:146-148(1997).
RN   [13]
RP   VARIANT ATR-X LEU-246.
RX   MEDLINE=20123062; PubMed=10660327;
RA   Fichera M., Romano C., Castiglia L., Failla P., Ruberto C., Amata S.,
RA   Greco D., Cardoso C., Fontes M., Ragusa A.;
RT   "New mutations in XNP/ATR-X gene: a further contribution to
RT   genotype/phenotype relationship in ATR/X syndrome.";
RL   Hum. Mutat. 12:214-214(1998).
RN   [14]
RP   VARIANT SHS LYS-1742.
RX   MEDLINE=99347960; PubMed=10417298;
RA   Lossi A.-M., Millan J.M., Villard L., Orellana C., Cardoso C.,
RA   Prieto F., Fontes M., Martinez F.;
RT   "Mutation of the XNP/ATR-X gene in a family with severe mental
RT   retardation, spastic paraplegia and skewed pattern of X inactivation:
RT   demonstration that the mutation is involved in the inactivation
RT   bias.";
RL   Am. J. Hum. Genet. 65:558-562(1999).
RN   [15]
RP   VARIANT CWS THR-2050.
RX   MEDLINE=99326061; PubMed=10398237;
RA   Abidi F., Schwartz C.E., Carpenter N.J., Villard L., Fontes M.,
RA   Curtis M.;
RT   "Carpenter-Waziri syndrome results from a mutation in XNP.";
RL   Am. J. Med. Genet. 85:249-251(1999).
RN   [16]
RP   VARIANTS ATR-X E-175; 178-V--K-198 DEL; S-190; P-219; L-246 AND C-249.
RX   MEDLINE=99219535; PubMed=10204841;
RA   Villard L., Bonino M.-C., Abidi F., Ragusa A., Belougne J.,
RA   Lossi A.-M., Seaver L., Bonnefont J.-P., Romano C., Fichera M.,
RA   Lacombe D., Hanauer A., Philip N., Schwartz C.E., Fontes M.;
RT   "Evaluation of a mutation screening strategy for sporadic cases of
RT   ATR-X syndrome.";
RL   J. Med. Genet. 36:183-186(1999).
RN   [17]
RP   VARIANTS ATR-X S-179; L-190; I-194; C-246; F-1552; S-1645 AND C-1847.
RX   MEDLINE=20451413; PubMed=10995512;
RA   Wada T., Kubota T., Fukushima Y., Saitoh S.;
RT   "Molecular genetic study of japanese patients with X-linked
RT   alpha-thalassemia/mental retardation syndrome (ATR-X).";
RL   Am. J. Med. Genet. 94:242-248(2000).
CC   -!- FUNCTION: COULD BE A GLOBAL TRANSCRIPTIONAL REGULATOR. MODIFIES
CC       GENE EXPRESSION BY AFFECTING CHROMATIN. MAY BE INVOLVED IN
CC       BRAIN DEVELOPMENT AND FACIAL MORPHOGENESIS.
CC   -!- SUBUNIT: PROBABLY BINDS EZH2. BINDS ANNEXIN V IN A CALCIUM AND
CC       PHOSPHATIDYLCHOLINE/PHOSPHATIDYLSERINE-DEPENDENT MANNER (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR. ASSOCIATED WITH PERICENTROMERIC
CC       HETEROCHROMATIN DURING INTERPHASE AND MITOSIS, PROBABLY BY
CC       INTERACTING WITH HP1.
CC   -!- ALTERNATIVE PRODUCTS: 5 ISOFORMS; 1, 2, 3, 4 (SHOWN HERE) AND 5;
CC       ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: UBIQUITOUS.
CC   -!- DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF X-LINKED ALPHA-
CC       THALASSEMIA/MENTAL RETARDATION SYNDROME (ALSO KNOWN AS ATR-X
CC       SYNDROME). ATR-X IS AN X-LINKED DISORDER COMPRISING SEVERE
CC       PSYCHOMOTOR RETARDATION, FACIAL DYSMORPHISM, UROGENITAL
CC       ABNORMALITIES, AND ALPHA-THALASSEMIA. AN ESSENTIAL PHENOTYPIC
CC       TRAIT ARE HEMOGLOBIN H ERYTHROCYTE INCLUSIONS.
CC   -!- DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF CARPENTER-WIZIRI
CC       SYNDROME (CWS), AN X-LINKED RECESSIVE CONDITION CHARACTERIZED BY
CC       MODERATE MENTAL RETARDATION, SHORT STATURE, BRACHYDACTYLY WITH
CC       EXCESSIVE SKIN CREASES, AND WIDENING OF THE KNUCKLES.
CC   -!- DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF JUBERG-MARSIDI SYNDROME
CC       (JM). JM IS A RARE X-LINKED RECESSIVE DISEASE CHARACTERIZED BY
CC       SEVERE MENTAL RETARDATION, GROWTH FAILURE, SENSORINEURAL DEAFNESS,
CC       MICROGENITALISM AND EARLY DEATH.
CC   -!- DISEASE: DEFECTS IN ATRX ARE A CAUSE OF SMITH-FINEMAN-MYERS
CC       SYNDROME (SFM). CLINICAL FEATURES INCLUDE SEVERE MENTAL
CC       RETARDATION, MICROCEPHALY, GROWTH FAILURE, FACIAL ANOMALIES AND
CC       BILATERAL CRYPTORCHIDISM. DUE TO THE CLINICAL OVERLAP WITH ATR-X
CC       SYNDROME, SOME PATIENTS ORIGINALLY DIAGNOSED AS HAVING SFM, MIGHT
CC       BE AFFECTED BY A VARIANT OF ATR-X SYNDROME WHICH LACK HEMOGLOBIN H
CC       INCLUSIONS.
CC   -!- DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF SUTHERLAND-HAAN X-LINKED
CC       MENTAL RETARDATION SYNDROME (SHS). IT IS CHARACTERIZED BY SEVERE
CC       MENTAL RETARDATION WITH SPASTIC PARAPLEGIA, MICROCEPHALY, SHORT
CC       STATURE AND CRYPTORCHIDISM.
CC   -!- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 PHD-TYPE ZINC FINGER.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U72937; AAB49970.2; -.
DR   EMBL; U72938; AAB49971.2; -.
DR   EMBL; U72935; AAB40698.1; -.
DR   EMBL; U72904; AAB40698.1; JOINED.

  Query Match             3.8%;  Score 208.5;  DB 1;  Length 2492;
  Best Local Similarity   21.3%;  Pred. No. 0.004;
  Matches  150;  Conservative  102;  Mismatches  286;  Indels  165;  Gaps   31;

Qy      62 RGYVALLQKKDPKFCSLSRIFHDQKKCDEHKASSSPFSV--------AKFRRWDCSKCLD 113
           : ::|| :  : :| ::  : : :|   |||  : |             : | ||
Db     404 KAHLALEEDLNSEFRAMDAV-NKEKNTKEHKVIDAKFETKARKGEKPCALEKKDISKSEA 462

Qy     114 KL----------------------------KTSDNGTAPRTLPAKQNGTSDGCSITFVR 144
           ||                            | ||    |:  ||    ||:   :  |
```

```
Db    463 KLSRKQVDSEHMHQNVPTEEQRTNKSTGGEHKKSDRKEEPQYEPA---NTSEDLDMDIVS 519

Qy    145 STFVPASVGSQ-----KVSPSTQSSQGKNADRSTLPKSVQEGNDSKCNAPSGKN--GAAE 197

Db    520 ---VPSSVPEDIFENLETAMEVQSSVDHQGDGSSGTEQEVESSSVKLNISSKDNRGGIKS 576

Qy    198 ANTDSPMKDLQGPAQNYDVAANVSEDNTSVDVGALPEVPQITWHIEVNGADQPPSTPKLS 257

Db    577 KTTAKVTKEL------YVKLTPVSLPNSPIKGADCQEVPQ-----DKDGYKSCGLNPKLE 625

Qy    258 EVVLKRNEDENG---KTEETLVAEQCNLTKDPNPMSGKERDQVAEQCNLTKDPKPVSGQK 314

Db    626 KCGLGQENSDNEHLVENEVSLLLEESDLRRSPRVKTTPLRRPT--------ETNPVTSNS 677

Qy    315 CEQICNEPCEEVVLKRSSKSKRKTDKKLMKKQQHSKKRTAQADVSDAKLCRRKPKKV--R 372

Db    678 DEE-CNETVKE--KQKLSVPVRKKDKR----------NSSDSAIDNPKPNKLPKS 719

Qy    373 LLSEIINANQVEDSRSDE-------VHRENAADPCEDDRSTIPVPMEVSMDIPVSNHTVG 425

Db    720 KQSETVDQN----SDSDEMLAILKGVSRMSHSSSSDTDSNEIHTNHKTLYDLKTQ---AG 772

Qy    426 EDGLKSSKNKTKRKYSDVVDD-----GSSLMNWLNGKKERTGSVHHT 467

Db    773 KD----DKGKRKRKSSTSGSDFDTKKGKSAKSSIISKKERQTQSESSNYDSELEKEIKSM 828

Qy    468 VAHPAGNLSNKKVTPTAS-TQHDDENDTENGLDTNMHKTDVCQHVSEISTQRCSSKGKTA 526

Db    829 SKIGAARTTKKRIPNTKDFDSSEDEKHSKKGMDNQGHK--------NLKTSQEGSSDDAE 880

Qy    527 GLSKGKTHSAASTKYGGESTRNGQNIHVLSAEDQCQMETENSVLSHSA-KVSPAEHDIQI 585

Db    881 RKQERETFSSAEGTVDKDTT-------IMELRDRLPKKQQASASTDGVDKLSGKE---QS 930

Qy    586 MSDLHEQSLPKKKKKQKLEVTREKQTMIDDIPMDIVE-LLAKNQHERQLMTETDCSDINR 644

Db    931 FTSLEVRKVAETKEKSKHLKTKTCKKVQDGL-SDIAEKELKKDQSD-------ETSEDDK 982

Qy    645 IQSKTTADDDCVIVAAKDGSDYASSVFDTNSQQKSLASQSTQK 687

Db    983 KQSKKGTEE------KKKPSDFKKKVIKMEQQYES-SSLGTEK 1018


RESULT   5
IGA4_HAEIN
ID   IGA4_HAEIN     STANDARD;      PRT;  1849 AA.
AC   P45386;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).
GN   IGA.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NHTI HK61;
RX   MEDLINE=92234949; PubMed=1373717;
RA   Poulsen K., Reinholdt J., Kilian M.;
RT   "A comparative genetic study of serologically distinct Haemophilus
RT   influenzae type 1 immunoglobulin A1 proteases.";
RL   J. Bacteriol. 174:2913-2921(1992).
CC   -!- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
CC       PRODUCING INTACT FC AND FAB FRAGMENTS.
CC   -!- CATALYTIC ACTIVITY: Cleavage of immunoglobulin a molecules at
CC       certain Pro-|-Xaa bonds in the hinge region. No small molecule
CC       substrates are known.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC       SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC       OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC       DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6 (SERINE PROTEASE).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M87491; AAA24968.1; -.
DR   MEROPS; S06.001; -.
DR   InterPro; IPR000710; IgA_S6.
DR   Pfam; PF02395; IGA1; 1.
DR   PRINTS; PR00921; IGASERPTASE.
KW   Hydrolase; Serine protease; Transmembrane; Zymogen; Signal.
FT   SIGNAL       1     25       POTENTIAL.
FT   CHAIN       26   1021       IMMUNOGLOBULIN A1 PROTEASE.
FT   PROPEP    1022   1849       HELPER PEPTIDE (POTENTIAL).
FT   ACT_SITE   299    299       PROBABLE.
SQ   SEQUENCE   1849 AA;  202957 MW;  79A7D018C7150AEA CRC64;


  Query Match             3.6%;  Score 200.5;  DB 1;  Length 1849;
  Best Local Similarity   18.4%;  Pred. No. 0.007;
  Matches  176;  Conservative  142;  Mismatches  403;  Indels  235;  Gaps   34;

Qy    130 KQNGTSDGCSITFVRSTFVPASVGSQKVSPSTQSSQGKNADRST-LPKSVQEGNDSKCNA 188

Db   1000 KRNQTVDTTNIT------TPNDI--QADAPSAQSNNEEIARVETPVPPP----------A 1041

Qy    189 PSGKNGAAEANTDSPMKDLQGPAQNYDVAANVSEDNTSVDVGALPEVPQITWHIEVNGAD 248

Db   1042 PATESAIASEQPETRPAETAQPAMEETNTANSTETAPKSDTATQTENP--------NSES 1093

Qy    249 QPPSTPKLSEVVLKRNEDENGKTEETLVAEQCNLTKDPNPMSGK--ERDQVAEQCNLTKD 306

Db   1094 VPSET---TEKVAENPPQEN----ETVAKNEQEAT-EPTPQNGEVAKEDQPTVEANTQTN 1145

Qy    307 PKPVSGQKCE-----QICNEPCEEVVLKRSSKSKRKT----DKKLMKKQQHSKKRTAQAD 357

Db   1146 EATQSEGKTEETQTAETKSEPTESVTVSENQPEKTVSQSTEDKVVVEKEEKAKVETEETQ 1205

Qy    358 VSDAKLCRRKPKKVRLLSEIINANQVEDSRSDEVHRENAADPCEDDRSTIPVPMEVSMDI 417

Db   1206 KAPQVTSKEPPKQA------------EPAPEEVPTDTNAEEAQALQQTQPTTV-AAAET 1251

Qy    418 PVSNHTVGEDGLKSSKNKTKRKYSDVVDDGSSLM-----NWLNGKKKRTGSVHHTVAHPA 472

Db   1252 TSPNSKPAEETQQPSEKTNAEPVTPVVSENTATQPTETEETAKVEKEKTQEVPQVASQES 1311

Qy    473 GNLSNKKVTPTASTQHDDENDTENGLDTNMHKTDVCQHVSEISTQRCSSKGKTAGLSKGK 532

Db   1312 PKQEQPAAKPQAQTKPQAEPARENVLTTKNVGEPQPQAQPQTQSTAVPTTGETAANSKPA 1371

Qy    533 THSAASTKYGGESTRNGQNIHVLSAEDQCQMETENSVLSHSAKVSPAEHDIQIMSDLHEQ 592

Db   1372 AKPQAQAKPQTEPAR--ENVSTVNTKEP-QSQT-------SATVSTEQPAKETSSNV-EQ 1420

Qy    593 SLPKKKKKQKLEVTREKQTMIDDIPMDIVELLAKNQHERQLMTETDCSDINRIQSKTTAD 652

Db   1421 PAPENSINTGSATTMTETAEKSDKPQ--METVTENDRQPEANTVADNSVANNSESS---- 1474

Qy    653 DDCVIVAAKDGSDYASSVFDTNSQQKSLASQSTQKELQGHLALTTQE--------SPHPQ 704

Db   1475 ------------------ESKSRRRRSVSQPKETSAEETTVASTQETTVDNSVSTPKPR 1515

Qy    705 NFQSTQEQQTHLRMEEMVTIAASSPLFSHHDDQYIAEAPTEHWGRKDAKKLTWEQFKATT 764

Db   1516 SRRTRRSVQTN----------SYEPV----------ELPTEN------------------ 1537
```

```
Qy     765 RNSPAATCGAQFRPGIQAVDLTSTHVMGSSSNYASRQPVIAPLDRYAERAVNQVHARNFP 824
                 |:    :|           | :| |: ||           |:  : ::|
Db    1538 ---------AENAENVQ----------SGNNVANSQP----------ALRNLTSKNTN 1566

Qy     825 STIATMEASKLCDRRNAGQVVLYPKESMPATHLLRMMDPSTLASFPNYGT----SSRNQM 880
           : ::   |       | |: |        : |: ::          | |      |   |
Db    1567 AVLSNAMAKAQFVALNVGKAV--------SQHISQL-------EMNNEGQYNVWISNTSM 1611

Qy     881 ESQLHNSQYAHNQYKGSTSTSYGSNLNGKIPLTFEDLSRHQLHDLHRPLRPHPRVGVLGS 940
               |  |:  ||:  :| |  : |     ::    :  ||  :   :|       |
Db    1612 -----NKNYSSEQYRRFSSKSTQTQLGWDQTIS----NNVQLGGVFTYVRNSNNFDKASS 1662

Qy     941 LLQKEIANWSENCGTQSGYKLGVSTGITSHQMNRKEH---------------FEALNSG 984
                 |:       : : ||:  |    | | : :                 :| | |
Db    1663 KNTLAQVNFYSKYYADNHWYLGIDLGYGKFQSNLQTNNNAKFARHTAQIGLTAGKAFNLG 1722

Qy     985 MFSAKWNALQLGSVSSSADFLSARNSIAQSWTRGKGKMVHPLD---RFVRQDICIT 1037
           |: |       |  |:|||  |:: |           |:|:     | : |:  |
Db    1723 NFAVKPTVGVRYSYLSNADFALAQDRI----------KVNPISVKTAFAQVDLSYT 1768
```

RESULT 6
ICAL_HUMAN
ID ICAL_HUMAN STANDARD; PRT; 708 AA.
AC P20810;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Calpain inhibitor (Calpastatin) (Sperm BS-17 component).
GN CAST.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91124109; PubMed=2577276;
RA Asada K., Ishino Y., Shimada M., Shimojo T., Endo M., Kimizuka F.,
RA Kato I., Maki M., Hatanaka M., Murachi T.;
RT "cDNA cloning of human calpastatin: sequence homology among human,
RT pig, and rabbit calpastatins.";
RL J. Enzym. Inhib. 3:49-56(1989).
RN [2]
RP SEQUENCE OF 1-283 FROM N.A., AND ALTERNATIVE SPLICING.
RC TISSUE=Lung;
RX MEDLINE=92235069; PubMed=1569094;
RA Lee W.J., Ma H., Takano E., Yang H.Q., Hatanaka M., Maki M.;
RT "Molecular diversity in amino-terminal domains of human calpastatin
RT by exon skipping.";
RL J. Biol. Chem. 267:8437-8442(1992).
RN [3]
RP SEQUENCE OF 101-317 FROM N.A.
RX MEDLINE=90037002; PubMed=2553724;
RA Maki M., Bagci H., Hamaguchi K., Ueda M., Murachi T., Hatanaka M.;
RT "Inhibition of calpain by a synthetic oligopeptide corresponding to
RT an exon of the human calpastatin gene.";
RL J. Biol. Chem. 264:18866-18869(1989).
RN [4]
RP SEQUENCE OF 125-283 FROM N.A.
RX MEDLINE=90165962; PubMed=2407243;
RA Uemori T., Shimojo T., Asada K., Asano T., Kimizuka F., Kato I.,
RA Maki M., Hatanaka M., Murachi T., Hanzawa H., Arata Y.;
RT "Characterization of a functional domain of human calpastatin.";
RL Biochem. Biophys. Res. Commun. 166:1485-1493(1990).
RN [5]
RP SEQUENCE OF 153-708 FROM N.A. (SHORT ISOFORM), AND REVISIONS.
RA Wang L.F.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
RN [6]
RP SEQUENCE OF 523-708 FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=95038502; PubMed=7951045;
RA Wang L.F., Wei S.G., Miao S.Y., Liu Q.Y., Koide S.S.;
RT "Calpastatin gene in human testis.";
RL Biochem. Mol. Biol. Int. 33:245-252(1994).
RN [7]
RP SEQUENCE OF 425-708 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=95221646; PubMed=7706496;
RA Despres N., Talbot G., Plouffe B., Boire G., Menard H.A.;
RT "Detection and expression of a cDNA clone that encodes a polypeptide
RT containing two inhibitory domains of human calpastatin and its
RT recognition by rheumatoid arthritis sera.";
RL J. Clin. Invest. 95:1891-1896(1995).
RN [8]
RP SEQUENCE OF 244-708 FROM N.A.
RC TISSUE=Placenta;
RA El-Amine M., Talbot G., Despres N., Asselin C., Boire G.,
RA Menard H.A.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DDBJ databases.
RN [9]
RP PHOSPHORYLATION.
RX MEDLINE=91139699; PubMed=1995645;
RA Adachi Y., Ishida-Takahashi A., Takahashi C., Takano E., Murachi T.,
RA Hatanaka M.;
RT "Phosphorylation and subcellular distribution of calpastatin in human
RT hematopoietic system cells.";
RL J. Biol. Chem. 266:3968-3972(1991).
CC -!- FUNCTION: SPECIFIC INHIBITION OF CALPAIN (CALCIUM-DEPENDENT
CC CYSTEINE PROTEASE). PLAYS A KEY ROLE IN POSTMORTEM TENDERIZATION
CC OF MEAT AND HAVE BEEN HYPOTHESIZED TO BE INVOLVED IN MUSCLE
CC PROTEIN DEGRADATION IN LIVING TISSUE.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- DOMAIN: HAS FOUR INHIBITORY DOMAINS.
CC -!- PTM: THE N-TERMINUS IS BLOCKED.
CC -!- PTM: PHOSPHORYLATED.
CC ----------

CC ----------
DR EMBL; D16217; BAA03747.1; -.
DR EMBL; D50827; BAA09438.1; -.
DR EMBL; M86258; AAB59398.1; -.
DR EMBL; M28230; AAA52066.1; -.
DR EMBL; M28227; AAA52066.1; JOINED.
DR EMBL; M28228; AAA52066.1; JOINED.
DR EMBL; M28229; AAA52066.1; JOINED.
DR EMBL; M33328; AAA52296.1; -.
DR EMBL; U26724; AAC50136.2; -.
DR EMBL; S73329; AAB32311.1; -.
DR EMBL; U38525; AAA80684.1; -.
DR EMBL; U31345; AAB60371.1; -.
DR EMBL; U31346; AAB60372.1; -.
DR PIR; A34428; A34428.
DR MIM; 114090; -.
DR InterPro; IPR001259; Calpain_inhib.
DR Pfam; PF00748; Calpain_inhib; 4.
KW Repeat; Thiol protease inhibitor; Alternative splicing;
KW Phosphorylation.

FT	DOMAIN	1	169 L.
FT	REPEAT	170	222 INHIBITORY DOMAIN 1.
FT	REPEAT	304	356 INHIBITORY DOMAIN 2.
FT	REPEAT	446	499 INHIBITORY DOMAIN 3.
FT	REPEAT	583	636 INHIBITORY DOMAIN 4.
FT	VARSPLIC	212	224 MISSING (IN SHORT ISOFORM).
FT	VARIANT	592	592 G -> E.
FT			/FTId=VAR_005298.
FT	CONFLICT	467	467 R -> L (IN REF. 5).
FT	CONFLICT	486	488 VKD -> GKE (IN REF. 5).
FT	CONFLICT	543	543 V -> L (IN REF. 5).

```
FT   CONFLICT    562    562       MISSING (IN REF. 8).
SQ   SEQUENCE   708 AA;  76484 MW;  CACD759C92841[3EA CRC64;


  Query Match             3.6%;  Score 197;  DB 1;  Length 708;
  Best Local Similarity   20.3%;  Pred. No. 0.00[1;
  Matches  155;  Conservative  109;  Mismatches  284;  Indels  216;  Gaps   35;

Qy      70 KKDPKFCSLS---RIFHDQK----KCDEH-------KA[SSPFSVAKFRRWDCSKCLDKL 115
           | :|:  | |    : |::|    |  ||       | :|   |     :   |:  ::
Db      32 KTEPEKKSQSTKLSVVHEKKSQEGKPKEHTEPKSLPKQ[SDTGSNDAHNKKAVSRSAEQ- 90

Qy     116 KTSDNGTAPRTLPAKQNGTSDGCSITFVRSTFVPASVG[QKVSPSTQSSQGKNADRSTLP 175
           : |:  | |:| |                          | | : |:  |  | ||  |:
Db      91 QPSEKSTEPKTKPQDM-----------------ISAG[ESVAGITAIS-GKPGDKKKEK 131

Qy     176 KSVQEG--NDSKCNAPSGKNGAAEANTDSPMKDLQGPAC[NYDVAANVSEDNTSVDVGALP 233
           ||:     :|| : ||||:|  :|  |  :  | || :          |:||:      |
Db     132 KSLTPAVPVESKPDKPSGKSG-MDAALDDLIDTLGGPEE[-------TEEENTTY---TGP 180

Qy     234 EV--PQITWHIEVNGADQPPSTPKLSEVVLKR---------------------------- 263
           ||  |  : :||  |  :    ||  |:: |:
Db     181 EVSDPMSSTYIEELGKREVTIPPKYRELLAKKEGITGPPADSSKPIGPDDAIDALSSDFT 240

Qy     264 --------NEDENGKTEETLVAEQCNLTKDPNPMSGKER--------DQVAEQCNL---T 304
                  : |  :: | | |:      :    |   |:|        ||  |  :     |
Db     241 CGSPTAAGKKTEKEESTEVLKAQSAGTVRSAAPPQEKKRKVEKDTMSDQALEALSASLGT 300

Qy     305 KDPKP-------------VSGQKCEQICNEPCEEVVLK-RSSKSKRKTDKKLMKKQQHS 349
           :  :|                :| |: | |  | :  : |   :  |  | |: : :
Db     301 RQAEPELDLRSIKEVDEAKAKEEKLEK-CGEDDETIPSEYRLKPATDKDGKPLLPEPEEK 359

Qy     350 KKRTAQADV-------SDAKLCRRKPKKVRLLSEIINAN[VEDSR-------SDEVHREN 395
           |  :::::          |   |: || |           : |:|:       |: | | :
Db     360 PKPRSESELIDELSEDFDRSECKEKPSK--------PTE[TEESKAAAPAPVSEAVSRTS 411

Qy     396 AADPCEDDRSTIPVPMEVSMDIPVSNHTVGEDGLKSSKN[TKRKYSDVVDDGSSLMNWLN 455
                 :|  | |  :   :|       :| :::  :   :| :|  :||  :|: :
Db     412 MC----SIQSAPPEPATLKGTVP-------DDAVEALAD[LGKKEAD-PEDGKPVMDKVK 459

Qy     456 GKKK-----RTGSVHHTV-----AHPAGNLSNKKVTPTA[TQH--------------DDE 491
           | |     : |    |:          :   | : |  | :                  :|
Db     460 EKAKEEDREKLGEKEETIPPDYRLEEVKDKDGKPLLPKE[KEQLPPMSEDFLLDALSEDF 519

Qy     492 NDTENGLDTNMHKTDVCQHVSEISTQRCSSKGKTAGLSK[KTHSAASTKYGGESTRNGQN 551
           :  :|          :  :||: :|             : |||   |   |:
Db     520 SGPQNASSLKFEDAKLAAAISEVVSQ----------------TPASTTQAGAPPRD--- 559

Qy     552 IHVLSAEDQCQMETENSVLSHS-AKVSPAEHDIQIMSDL[EQSLPKKKKKQKLEVTREKQ 610
             :::    ::     || |  :  |   : : | |       | |:| | |  |:|
Db     560 ----TSQSDKDLDDALDKLSDSLGQRQPDPDENKPMGD-------KVKEKAKAE-HRDKL 607

Qy     611 TMIDD-IPMDIVELLAKNQHERQLMTETDCSDINRIQSK[TADDDCVIVAAKDGSDYASS 669
             || || :   ||  |  :: :   |  |:      ||  |||   | |      |   |
Db     608 GERDDTIPPEYRHLLDDNGQDKPVKPPTKKSE----DSK[PADDQDPIDALSGDLDSCPS 663

Qy     670 VFDTNSQQ-----KSLASQSTQKELQGHL---ALTTQES[HPQN 705
             :|:        |  || |   :  |     | ||:|:  |::
Db     664 TTETSQNTAKDKCKKAASSSKAPKNGGKAKDSAKTTEET[KPKD 707


RESULT   7
MAPB_HUMAN
ID   MAPB_HUMAN      STANDARD;      PRT;  2468 AA.
AC   P46821;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Microtubule-associated protein 1B (MAP 1B) [Contains: MAP1 light chain
DE   LC1].
GN   MAP1B.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Fetal brain;
RX   MEDLINE=95104835; PubMed=7806212;
RA   Lien L.L., Feener C., Fischbach N., Kunkel L.M.;
RT   "Cloning of human microtubule-associated protein 1B and the
RT   identification of a related gene on chromosome 15.";
RL   Genomics 22:273-280(1994).
CC   -!- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN.
CC       PHOSPHORYLATED MAP1B MAY PLAY A ROLE IN THE CYTOSKELETAL CHANGES
CC       THAT ACCOMPANY NEURITE EXTENSION. POSSIBLY MAP1B BINDS TO AT LEAST
CC       TWO TUBULIN SUBUNITS IN THE POLYMER, AND THIS BRIDGING OF SUBUNITS
CC       MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN
CC       STABILIZING MICROTUBULES.
CC   -!- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE
CC       WITH MAP1A AND MAP1B PROTEINS.
CC   -!- DOMAIN: Has a highly basic region with many copies of the sequence
CC       KKEE and KKEI/V, repeated but not at fixed intervals, which is
CC       responsible for the binding of MAP1B to microtubules.
CC   -!- PTM: LC1 is coexpressed with MAP1B. It is a polypeptide generated
CC       from MAP1B by proteolytic processing. It is free to associate with
CC       both MAP1A and MAP1B. It interacts with the amino-terminal region
CC       of MAP1B (By similarity).
CC   -!- SIMILARITY: TO MAP1A.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L06237; AAA18904.1; -.
DR   MIM; 157129; -.
DR   InterPro; IPR000102; MAP1B_neuraxin.
DR   Pfam; PF00414; MAP1B_neuraxin; 10.
DR   PROSITE; PS00230; MAP1B_NEURAXIN; 6.
KW   Microtubules; Repeat; Phosphorylation.
FT   CHAIN         ?   2468       MAP1 LIGHT CHAIN LC1.
FT   REPEAT     1878   1894       MAP1B 1.
FT   REPEAT     1895   1911       MAP1B 2.
FT   REPEAT     1912   1928       MAP1B 3.
FT   REPEAT     1929   1945       MAP1B 4.
FT   REPEAT     1946   1962       MAP1B 5.
FT   REPEAT     1963   1979       MAP1B 6.
FT   REPEAT     1997   2013       MAP1B 7.
FT   REPEAT     2014   2030       MAP1B 8.
FT   REPEAT     2031   2047       MAP1B 9.
FT   REPEAT     2048   2064       MAP1B 10.
FT   DOMAIN      589    790       LYS-RICH (HIGHLY BASIC, CONTAINS MANY
FT                                KKEE AND KKEI/V REPEATS).
SQ   SEQUENCE   2468 AA;  270618 MW;  540839CBDF09D461 CRC64;


  Query Match             3.6%;  Score 196.5;  DB 1;  Length 2468;
  Best Local Similarity   18.4%;  Pred. No. 0.016;
  Matches  204;  Conservative  170;  Mismatches  448;  Indels  289;  Gaps   44;

Qy     167 KNADRSTLPKSVQEGNDSKCNAPSGKNGAAEANTDS--PMKDLQGPAQNYDVAANVSEDN 224
           |: |    | : |:    :   |  |    :   ||   :|    |  :  |     |
Db     506 KHLDFLKQPLATQKDLTGQVPTPVVKQTKLKQRADSRESLKPAAKPLPSKSVRKESKE-- 563

Qy     225 TSVDVGALPEVPQITWHIEVNGADQPPSTPKLSEVVLKRNEDENGKTEETLVAEQCNLTK 284
                   | |::|   :||  ::||      :|::|:::    :|: ::  ::    :
Db     564 ---------ETPEVT---KVNHVEKPPKVESKEKVMVKKDKPVKTETKPSVTEKEVPSKE 611

Qy     285 DPNPMSGKERDQVAEQCNLTKDPKPVSG-----QKCEQICNEPCEEVVLKRSSKSKRKTD 339
           :|:|:    : :|||:     | || :      :|  ::  |  :|   |   :  :| |
Db     612 EPSPV----KAEVAEK--QATDVKPKAAKEKTVKKETKVKPEDKKEEKEKPKKEVAKKED 665
```

```
Qy    340 KKLMKKQQHSKKRTAQADV------SDAKLCRRKPKKVRLLSEIINANQVEDSRSDEVHR 393
Db    666 KTPIKKEEKPKKEEVKKEVKKEIKKEEKKEPKKEVKKETPPKEV--KKEVKKEEKKEVKK 723

Qy    394 ENAAD-------PCEDDRSTIPV-----PMEVSMDIPVSNHTVGEDGLKSSKNKTKRKYS 441
Db    724 EEKEPKKEIKKLPKDAKKSSTPLSEAKKPAALKPKVPKKEESVKKDSVAAGKPKEKGKIK 783

Qy    442 DVVDDGSSLMNWLNGKKKRTGSVHHTVAHPAGNLS---NKKVTPTASTQHDDENDTENGL 498
Db    784 VIKKEGKAAE--AVAAAVGTGATTAAVMAAAGIAAIGPAKELEAERSLMSSPEDLTKDFE 841

Qy    499 DTNMHKTDVCQHVSE--------------------ISTQRCSSKGKTAGLSKGKTHSAA 537
Db    842 ELKAEEVDVTKDIKPQLELIEDEEKLKETEPVEAYVIQKEREVTKGPAESPDEGIT---- 897

Qy    538 STKYGGESTRNGQNIHVL---SAEDQCQMETENSVLSHSAKV----------------SP 578
Db    898 TTEGEGECEQTPEELEPVEKQGVDDIEKFEDEGAGFEESSETGDYEEKAETEEAEEPEED 957

Qy    579 AEHDIQIMSDLHEQSLPKKKKKQKLEV-TREKQTMI----DDIPMDIVELLAK---NQHE 630
Db    958 GEEHVCVSASKHSPTEDEESAKAEADAYIREKRESVASGDDRAEEDMDEAIEKGEAEQSE 1017

Qy    631 RQLMTETDCSDINRIQ---SKTTADDDCVIV------------------------AAKD 662
Db   1018 EEADEEDKAEDAREEEYEPEKMEAEDYVMAVVDKAAEAGGAEEQYGFLTTPTKQLGAQSP 1077

Qy    663 GSDYASSVFDT----NSQQKSLASQSTQKE------------LQGHLALTTQESP----- 701
Db   1078 GREPASSIHDETLPGGSESEATASDEENREDQPEEFTATSGYTQSTIEISSEPTPMDEMS 1137

Qy    702 HPQNFQS--TQEQQTHLRMEEMVTIAA-SSPLFSHH----------------------D 735
Db   1138 TPRDVMSDETNNEETESPSQEFVNITKYESSLYSQEYSKPADVTPLNGFSEGSKTDATDG 1197

Qy    736 DQYIAEAPT---------EHWGRKDAKKLTWEQFKATTRNSPAATCGAQFRPGIQAVDLT 786
Db   1198 KDYNASASTISPPSSMEEDKFSRSALRDAYCSEVKASTT--------LDIKDSISAV--S 1247

Qy    787 STHVMGSSSNYASRQPVIAPLDR--YAERAVNQVHARNFPSTIATMEASKLCDRRNAGQV 844
Db   1248 SEKVSPSKSPSLSPSPP-SPLEKTPLGERSVNFSLTPNEIKVSAEAEVAPVSP--EVTQE 1304

Qy    845 VLYPKESMPATHLLRMMDPSTLASFPNYGTSSRNQMESQLHNSQYAHNQYKGSTSTSYGS 904
Db   1305 VVEEHCASPEDKTLEVVSP----------------SQSVTGSAGHTPYYQSPTDEKSS 1346

Qy    905 NLNGK-------IPLTFE------------------DLSRHQLHDLHRPLRPHPRVG- 936
Db   1347 HLPTEVIEKPPAVPVSFEFSDAKDENERASVSPMDEPVPDSESPIEKVLSPLRSPPLIGS 1406

Qy    937 --VLGSLLQKE--------IANWSENCGTQ-SGYKLGVSTGITSHQMNR---------- 974
Db   1407 ESAYESFLSADDKASGRGAESPFEEKSGKQGSPDQVSPVSEMTSTSLYQDKQEGKSTDFA 1466

Qy    975 --KEHF---------EALNSGMFSAKWNALQLGSVSSS----ADFLSARNSIAQSWTRG- 1018
Db   1467 PIKEDFGQEKKTDDVEAMSSQPALA-LDERKLGDVSPTQIDVSQFGSFKEDTKMSISEGT 1525

Qy   1019 -KGKMVHPLDRFVRQDICITNKNPADFTTIS 1048
Db   1526 VSDKSATPVDEGVAEDTYSHMEGVASVSTAS 1556


RESULT   8
ANK2_HUMAN
ID   ANK2_HUMAN     STANDARD;      PRT;  3924 AA.
AC   Q01484; Q01485;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Ankyrin 2 (Brain ankyrin) (Ankyrin B) (Ankyrin, nonerythroid).
GN   ANK2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
RC   TISSUE=Brain stem;
RX   MEDLINE=91302466; PubMed=1830053;
RA   Otto E., Kunimoto M., McLaughlin T., Bennett V.;
RT   "Isolation and characterization of cDNAs encoding human brain
RT   ankyrins reveal a family of alternatively spliced genes.";
RL   J. Cell Biol. 114:241-253(1991).
RN   [2]
RP   REVISIONS.
RA   Carpenter S.;
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Brain stem;
RX   MEDLINE=94075409; PubMed=8253844;
RA   Chan W., Kordeli E., Bennett V.;
RT   "440-kD ankyrinB: structure of the major developmentally regulated
RT   domain and selective localization in unmyelinated axons.";
RL   J. Cell Biol. 123:1463-1473(1993).
RN   [4]
RP   SEQUENCE OF 463-495 FROM N.A.
RX   MEDLINE=92009921; PubMed=1833308;
RA   Tse W.T., Menninger J.C., Yang-Feng T.L., Francke U., Sahr K.E.,
RA   Lux S.E., Ward D.C., Forget B.G.;
RT   "Isolation and chromosomal localization of a novel nonerythroid
RT   ankyrin gene.";
RL   Genomics 10:858-866(1991).
CC   -!- FUNCTION: Attach integral membrane proteins to cytoskeletal
CC       elements. Also bind to cytoskeletal proteins.
CC   -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1 (SHOWN HERE), 2 AND 3; ARE
CC       PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: PLASMA MEMBRANE OF NEURONS AS WELL AS GLIAL
CC       CELLS THROUGHOUT THE BRAIN.
CC   -!- PTM: PHOSPHORYLATED AT MULTIPLE SITES BY DIFFERENT PROTEIN KINASES
CC       AND EACH PHOSPHORYLATION EVENT REGULATES THE PROTEIN'S STRUCTURE
CC       AND FUNCTION (POTENTIAL).
CC   -!- SIMILARITY: CONTAINS 23 ANK REPEATS.
CC   -!- SIMILARITY: CONTAINS 1 DEATH DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X56957; CAA40278.1; -.
DR   EMBL; X56958; CAA40279.2; -.
DR   EMBL; Z26634; CAB42644.1; -.
DR   EMBL; M37123; AAA62828.1; -.
DR   PIR; S14533; S14533.
DR   PIR; A39643; A39643.
DR   PIR; B39643; B39643.
DR   PIR; S14569; S14569.
DR   HSSP; Q00421; 1AWC.
DR   MIM; 106410; -.
DR   InterPro; IPR002110; ANK.
DR   InterPro; IPR000488; Death.
DR   InterPro; IPR000906; ZU5.
DR   Pfam; PF00023; ank; 24.
DR   Pfam; PF00531; death; 1.
DR   Pfam; PF00791; ZU5; 1.
DR   PRINTS; PR01415; ANKYRIN.
DR   SMART; SM00248; ANK; 21.
DR   SMART; SM00005; DEATH; 1.
DR   SMART; SM00218; ZU5; 1.
```

```
DR   PROSITE; PS50088; ANK_REPEAT; 20.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
DR   PROSITE; PS50017; DEATH_DOMAIN; 1.
KW   Cytoskeleton; Alternative splicing; Repeat; ANK repeat;
KW   Phosphorylation.
FT   REPEAT       63     92       ANK 1.
FT   REPEAT       96    125       ANK 2.
FT   REPEAT      129    158       ANK 3.
FT   REPEAT      162    191       ANK 4.
FT   REPEAT      193    220       ANK 5.
FT   REPEAT      232    261       ANK 6.
FT   REPEAT      265    294       ANK 7.
FT   REPEAT      298    327       ANK 8.
FT   REPEAT      331    360       ANK 9.
FT   REPEAT      364    393       ANK 10.
FT   REPEAT      397    426       ANK 11.
FT   REPEAT      430    459       ANK 12.
FT   REPEAT      463    492       ANK 13.
FT   REPEAT      496    525       ANK 14.
FT   REPEAT      529    558       ANK 15.
FT   REPEAT      562    591       ANK 16.
FT   REPEAT      595    624       ANK 17.
FT   REPEAT      628    657       ANK 18.
FT   REPEAT      661    690       ANK 19.
FT   REPEAT      694    723       ANK 20.
FT   REPEAT      727    756       ANK 21.
FT   REPEAT      760    789       ANK 22.
FT   REPEAT      793    822       ANK 23.
FT   DOMAIN     1773   1950       REPEAT-RICH REGION.
FT   REPEAT     1773   1784       REPEAT A.
FT   REPEAT     1785   1796       REPEAT A.
FT   REPEAT     1797   1808       REPEAT A.
FT   REPEAT     1809   1820       REPEAT A.
FT   REPEAT     1821   1832       REPEAT A.
FT   REPEAT     1833   1844       REPEAT A.
FT   REPEAT     1845   1856       REPEAT A.
FT   REPEAT     1857   1867       REPEAT A (APPROXIMATE).
FT   REPEAT     1868   1879       REPEAT A.
FT   REPEAT     1880   1891       REPEAT A.
FT   REPEAT     1892   1902       REPEAT A (APPROXIMATE).
FT   REPEAT     1903   1914       REPEAT A.
FT   REPEAT     1915   1926       REPEAT A.
FT   REPEAT     1927   1938       REPEAT A.
FT   REPEAT     1939   1950       REPEAT A.
FT   DOMAIN     3536   3620       DEATH.
FT   VARSPLIC   1039   1039       Q -> QFLGKLHLPTAPPPLNEGESLVSRILQLGPPGTK
FT                                (IN ISOFORM 2).
FT   VARSPLIC   1444   3528       MISSING (IN ISOFORM 2 AND ISOFORM 3).
FT   CONFLICT    475    476       GQ -> PE (IN REF. 4).
FT   CONFLICT    971    971       I -> S (IN REF. 1).
FT   CONFLICT   3581   3582       QY -> HA (IN REF. 1).
FT   CONFLICT   3586   3586       I -> Y (IN REF. 1).
SQ   SEQUENCE   3924 AA;  430337 MW;  52AC496C428E29D2 CRC64;

  Query Match             3.5%;  Score 194.5;  DB 1;  Length 3924;
  Best Local Similarity   18.7%;  Pred. No. 0.038;
  Matches  199;  Conservative  167;  Mismatches  391;  Indels  309;  Gaps   46;

Qy      28 VAPVLELTATPRQDAAAEAGVDEPAQH--------QCEHPSIRGYVALLQKKDPKFCSLS 79
           | |     :||:     |   :: :::         : | | :   :  :  : : :
Db    2517 VTPKTTDVSTPKPAVIHECAEEDDSENGEKKRFTPEEEMPKMVTKIKMFDELEQE-AKQK 2575

Qy      80 RIFHDQKKCDEHKASSSPFSVAKFRRWDCSKCLDKLKTSDNGTAPRTLPAKQNGTSDGCS 139
           | :  : | :|  :|| | :       |||  :|: | : :|     :|
Db    2576 RDYKKEPKQEESSSSSDPDA-------DCSVDVDEPKHTGSGEDESGVP---------- 2617

Qy     140 ITFVRSTFVPASVGSQKVSPSTQSSQ-----GKNADRSTLPKSVQEGNDSKCNAPSGKNG 194
                 |  :  |:||| |::|        | ||    |: |
Db    2618 --------VLVTSESRKVSSSSESEPELAQLKKGADSGLLPEPV---------------- 2653

Qy     195 AAEANTDSPMKDLQGPAQNYDVAANVSEDNTSVDVGALPIVPQITWHIEVNGADQPPSTP 254
                 ||:                    :|:|  : ||  |           ||  :
```

```
Db    2654 -IRVQPPSPLP------------------SSMDSNSSPEEVQF----------QPVVS- 2682

Qy     255 KLSEVVLKRNED---ENGKTEETLVAEQCNLTKDPNPMSGKERDQVAEQCNL-TKDPKPV 310
             :   | |||   | ||:||   :|  :| :| : :| :  |:  :: |  |  ||
Db    2683 --KQYTFKMNEDTQEEPGKSEEEKDSES-HLAEDRHAVSTEAEDRSYDKLNRDTDQPKIC 2739

Qy     311 SGQKCEQIC----------------NEPCEEVVLKRSSKSKRKTDKKLMKKQQ--HSKK 351
           |  || :                  ::  |: |: : | : : | :|  : ::   |:
Db    2740 DGHGCEAMSPSSSARPVSSGLQSPTGDDVDEQPVIYKESLALQGTHEKDTEGEELDVSRA 2799

Qy     352 RTAQADV------SDAKLCRRKPKKVRLLSEIINA----NQVEDSRSDEVHRENAADPCE 401
           : |||       | : |      : : | | |:|     : | :::||   ||    |
Db    2800 ESPQADCPSESFSSSSSLPHCLVSEGKELDEDISATSSIQKTEVTKTDETF-ENLPKDCP 2858

Qy     402 DDRSTIPVPME-VSMDIPVSN----------------HTVGEDGLKSSKNKTK------ 437
             |:|    :  |||:|||:                  :   :    | ::||:
Db    2859 SQDSSITTQTDRFSMDVPVSDLAENDEIYDPQITSPYENVPSQSFFSSEESKTQTDANHT 2918

Qy     438 ------RKYS--------DVVDDGSSLMNWLNGKK-------KRTGSVHHTVAHPAGNLS 476
                 ||        |||   ||    |: :        |    :  |:     :
Db    2919 TSFHSSEVYSVTITSPVEDVVVASSSSGTVLSKESNFEGQDIKMESQLESTLWEMQSDSV 2978

Qy     477 NKKVTPTAS---------------TQHDDENDT-------ENGLDTNMHKTDVCQHVSEI 514
           :   || |                |: | ::|:       :   : : : :  | :  :
Db    2979 SSSFEPTMSATTTVVGEQISKVIITKTDVDSDSWSEIREDDEAFEARVKEEE--QKIFGL 3036

Qy     515 STQRCSSKGKTAGLSKGKTHSAASTKYGGESTRNGQNIHVLSAEDQCQMETENSV-LSHS 573
             |  |:| |   :  :| :   |         : ||  :        :|    :: ::
Db    3037 MVDR-QSQGTTPDTTPARTPTEEGT------PTSEQNPFLFQEGKLFEMTRSGAIDMTKR 3089

Qy     574 AKVSPAEHDIQIMSDLHEQSLPKKKKKQKLEVTREKQTMIDDIPMDIVELLAKNQHERQL 633
           :   : |  ||  :  |::|          |  ;|  |  | :|:
Db    3090 SYADESFHFFQIGQESREETLS--------EDVKEGATGADPLPL--------------- 3126

Qy     634 MTETDCSDINRIQSKTTADDDCVIV---AAKDGSDYASSVFDTNSQQKSLASQSTQKELQ 690
             ||    :   :|| | ||:  ::   :::  :  :|    |||     | ||:  :
Db    3127 --ETSAESLALSESKETVDDEADLLPDSVSEEVEEIPASDAQLNSQMG--ISASTETPTK 3182

Qy     691 GHLALTTQESPHPQ-------------NFQSTQEQQTHLRMEEMVTIAAS--------SP 729
             ::: |:: |  |              :   : |    ::: :  |:  |          ||
Db    3183 EAVSVGTKDLPTVQTGDIPPLSGVKQISCPDSSEPAVQVQL-DFSTLTRSVYSDRGDDSP 3241

Qy     730 LFSHHDDQYIAEAPTEHWGRKDAKKLTWEQFKATTRNSPAAT---CGAQFRPGIQAVDLT 786
             |  : : : | ||      :    |  : |   |  | :|     |::   :    |:
Db    3242 DSSPEEQKSVIEIPT---APMENVPFTESKSKIPVRTMPTSTPAPPSAEYESSVSEDFLS 3298

Qy     787 STHVMGSSSNYA------SRQPVIAPLDRYAERAVNQVHARNFPSTIATM---EASKLCD 837
           |        | |       |: ||  || |  |:::: :   :   |:|      ||  |
Db    3299 SV----DEENKADEAKPKSKLPVKVPLQR-VEQQLSDLDT-SVQKTVAPQGQDMASIAPD 3352

Qy     838 RRNAGQVVLYPKESMPATHLLRMMDPSTLASFPNYGTSSRNQMESQLHNSQYAH----NQ 893
             |:  :                  | |:| |       :|:  |:: .: : |
Db    3353 NRSKSE----------------SDASSLDSKTKCPVKTRSYTETETESRERAEELELES 3395

Qy     894 YKGST-----------STSYGSNLNGKIPLTFEDLSRHQLHDLHR 927
           :|:|            | |  |:  |    | |  |:    ||:|
Db    3396 EEGATRPKILTSRLPVKSRSTTSSCRGGTSPTKE--SKEHFFDLYR 3439
```

```
RESULT   9
YFI6_YEAST
ID   YFI6_YEAST     STANDARD;      PRT;  1233 AA.
AC   P43597;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   Hypothetical 137.7 kDa protein in UGS1-FAB1 intergenic region.
GN   YFR016C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RX   MEDLINE=95400292; PubMed=7670463;
RA   Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,
RA   Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,
RA   Yamazaki M., Tashiro H., Eki T.;
RT   "Analysis of the nucleotide sequence of chromosome VI from
RT   Saccharomyces cerevisiae.";
RL   Nat. Genet. 10:261-268(1995).
CC   -!- SIMILARITY: SOME, TO MAMMALIAN NEUROFILAMENT TRIPLET M PROTEIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D50617; BAA09255.1; -.
DR   SGD; S0001912; YFR016C.
KW   Hypothetical protein.
SQ   SEQUENCE   1233 AA;  137697 MW;  C8A7CD2C6F0892F6 CRC64;


  Query Match             3.5%;  Score 191;  DB 1;  Length 1233;
  Best Local Similarity   18.9%;  Pred. No. 0.013;
  Matches  198;  Conservative  147;  Mismatches  385;  Indels  318;  Gaps   48;

Qy      50 EPAQHQCEHFSIRGYVALLQKKDPKFCSLSRIFHDQKKCDEHKASSSPFSVAKFRRWDCS 109
           |  | : :   |   |    :|: :      :|  : | ||
Db     360 EKQQTESDRDGISPSVLAKNQKETEIGKEDHVFEQKDKEDE------------------- 400

Qy     110 KCLDKLKT--------------SDNGTAPRTL----------FAK---------QNGTSD 136
           ||  :|                ||:   | |          | |          |:||:|
Db     401 KCRKELSVNHENNMSHNFNAAGSDSIIPPETERETYDDETMGPTKRISDNEKNLQHGTND 460

Qy     137 GCSITFVR-----------STFVPASVGSQKVSPSTQSSQGKNADRSTLPKSVQEGNDSK 185
             |:   :           |||  : |  : |:   | :   |    |  :|  :| :
Db     461 -ISVEVEKEEEEEEEEEENSTF--SKVKKENVT-GEQEAVRNNEVSGTEEESTSKGEE-- 514

Qy     186 CNAPSGKNGAAEANTDSPMKDLQGPAQNYDVAANVSEDNTSVDVGALPEVP-QITWHIEV 244
           |    :||   | : :::| |     :| :|:||  ::  |  | | |     ||
Db     515 --IMGGDEKQSEAGEKSSIIEIEGSAN----SAKISKDNLVLEDEA--EAPTQENKPTEV 566

Qy     245 NGADQPPSTPKLSEVVLKRNEDENGKTEETLVAEQCNLTKDPNPMSGKERDQVAEQCNLT 304
           |   | |:    :::          |: :: |   : |:          ||  ||  |
Db     567 VGEIDIPDAPRDDVEIVE-------AVEKNIIPEDLEVAKE---------DQEGEQVKLD 610

Qy     305 KDPKPVSGQKCEQICNEPCEEVVLKRSSKSKRKTDKKLMKKQQHSKKRTAQADVSDAKLC 364
           :  | :   |       |   | : |:   :   :::|  |:   : : :|:    :
Db     611 EPVKAMKDDKIAMRGAESISEDMKKKQEGTAELSNEKAKKEVDETARESAEGVEVEKSKT 670

Qy     365 RRKPKKVRLLSEIINANQVED----SRSDEVHRENAADPCEDDRSTIPVPMEVSMD---- 416
              || |:  :    : : ||      |  || :|:   | || :  |   :| | :
Db     671 PESPKVVKRCT----SGRPEDLQINERDPEVLKEDVRVPDEDVKPEIATTIENSEEEDPK 726

Qy     417 ---IPVSNH-------TVGEDGLKSS-KNKTKRKYSDVVDDGSSL----MNWLNGKKKRT 461
             : :|           :|: |  :| | : | |  ::  :|  :      |  :|:
Db     727 SQRVQISTEQAETTQKDMGDVGSTTSFKEEEKPKRFEITQEGDKITGKDTNHEHGEATEA 786

Qy     462 GSVHHTVAHPAGNLSNKKVTPTA-STQHDDENDTENGLDTNMHKTDVCQHVSEISTQRCS 520
           |  :: |   |  : | : |:: | : | | | |     |  ||:  :  :|:
Db     787 AS-ENSKASDVGT-AEKYIEPSSESVKKDTEEDAE---VENSEKTEFIKVKAELENLDAP 841

Qy     521 SKGK-TAGLSKGKTHSAASTKYGGESTRNGQNIHVLSAEDQCQMETENSVLSHSAKVSPA 579
           : : || |:|                    ::: | : ||    | |||  :   ||
Db     842 KEAEVTAELNK----------------ENEDVEVDTEED---AEVENSEKTEFIKVK-- 879

Qy     580 EHDIQIMSDLHEQSLPKKKKKQKLEVTREKQTMIDDIPMDIVELLAKNQHERQLMTETDC 639
                ::|     ||     : ||| |     :|     ||: | :: :     || |
Db     880 -------AELGNLDAPK-----EAEVTAELNKENED-----VEVAATSKED----IETKC 918

Qy     640 SDINRIQSKTTADD----------DCVIVAAKD----GSDYASSVFDTNSQQ------- 677
           |:     ::|  :|           |   |  :|     |  | :: |    |
Db     919 SE----PAETPIEDGTCTEAEVSKKDAEAVTKEDENMENSKIAEALKDVTGDQEIDDINI 974

Qy     678 -------------------------KSLASQSTQKELQ-GHLALTTQESPHPQNFQSTQE 711
                                     | |  : |:||     | :     :      : :|
Db     975 SDEFQRTVELPELEKQDIKDNKGEDKELEVEETEKETSLPDLVVEENITEEKNEIKQEEE 1034

Qy     712 QQTHLRMEEMVTIAASSPLFSHHDDQYIAEAPTEHWGRKDAKKLT-------------- 756
           : : |   |  :|:  :|   ::|:    :  |    |:: || :
Db    1035 EVSQLDFNETESISKEAP---NNDENGFEDQST----RENPKKASADDIFKDILDETNEF 1087

Qy     757 WEQFKATTRNSPAATCGAQFRPGIQAVDLTSTHVMGSSSNYASRQPVIAPLDRYAERAVN 816
           || |    :   |         : | | |:     |  |    | |
Db    1088 LEQLKIVDDSELNA-----LLQSLDAKDSTTQTTEQSKKNNDKPQDV------------ 1129

Qy     817 QVHARNFPSTIATMEASKLCDRRNAGQVVLYPKESMPATHLLRMMDPSTLASFPNYGTSS 876
                 | | |  || ::    | :|   :    |::    | |
Db    1130 ----------ITTSEIRKLNEKE---PVYIYTSLAGGGFHMI----PRT---------- 1161

Qy     877 RNQMESQLHNSQYAHNQYKGSTSTSYGSNLNGKIPLTFEDLSRHQLHDLHRPLRPHPRVG 936
           |::                  ||   :|   :|| |: ||      :  :  :   :
Db    1162 -NRL-----------------STILTAN---RIPFTYRDLGTDD--EARKVWKTFSKGR 1197

Qy     937 VLGSLLQKE---IANWSENCGTQSGYKL 961
           |  :::     | || |       |||
Db    1198 SLPGVVRGHNDLIGNWEEIEEANEDYKL 1225


RESULT  10
RP1_HUMAN
ID   RP1_HUMAN      STANDARD;      PRT;  2156 AA.
AC   P56715;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Oxygen-regulated protein 1 (Retinitis pigmentosa RP1 protein)
DE   (Retinitis pigmentosa 1 protein).
GN   RP1 OR ORP1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND VARIANTS HIS-872; TYR-985; THR-1670; PRO-1691
RP   AND TYR-2033.
RC   TISSUE=Retina;
RX   MEDLINE=99318096; PubMed=10391212;
RA   Sullivan L.S., Heckenlively J.R., Bowne S.J., Zuo J., Hide W.A.,
RA   Gal A., Denton M., Inglehearn C.F., Blanton S.H., Daiger S.P.;
RT   "Mutations in a novel retina-specific gene cause autosomal dominant
RT   retinitis pigmentosa.";
RL   Nat. Genet. 22:255-259(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Retina;
RX   MEDLINE=99318095; PubMed=10391211;
RA   Pierce E.A., Quinn T., Meehan T., McGee T.L., Berson E.L., Dryja T.P.;
RT   "Mutations in a gene encoding a new oxygen-regulated photoreceptor
RT   protein cause dominant retinitis pigmentosa.";
RL   Nat. Genet. 22:248-254(1999).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99330563; PubMed=10401003;
RA   Guillonneau X., Piriev N.I., Danciger M., Kozak C.A., Cideciyan A.V.,
RA   Jacobson S.G., Farber D.B.;
RT   "A nonsense mutation in a novel gene is associated with retinitis
RT   pigmentosa in a family linked to the RP1 locus.";
RL   Hum. Mol. Genet. 8:1541-1546(1999).
CC   -!- FUNCTION: COULD HAVE A ROLE IN THE DIFFERENTIATION OF
CC       PHOTORECEPTOR CELLS.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN RETINA. NOT EXPRESSED IN HEART,
CC       BRAIN, PLACENTA, LUNG, LIVER, SKELETAL MUSCLE, KIDNEY, SPLEEN AND
```

```
CC       PANCREAS.
CC   -!- DISEASE: DEFECTS IN RP1 CAUSE RETINITIS PIGMENTOSA FORM 1 (RP1); A
CC       DISEASE CHARACTERIZED BY CONSTRICTION OF THE VISUAL FIELDS, NIGHT
CC       BLINDNESS, AND FUNDUS CHANGES. THE DISEASE SEEMS TO BE ASSOCIATED
CC       WITH TRUNCATED (STOP OR FRAMESHIFT MUTATIONS) FORMS OF THE
CC       PROTEIN.
CC   -!- SIMILARITY: CONTAINS 2 DOUBLECORTIN DOMAINS.
CC   -!- DATABASE: NAME=RetNet;
CC       NOTE=Retinal information network;
CC       WWW="http://www.sph.uth.tmc.edu/retnet/".
CC   -!- DATABASE: NAME=Mutations of the RP1 gene;
CC       NOTE=Retina International's Scientific Newsletter;
CC       WWW="http://www.retina-international.com/sci-news/rp1mut.htm".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF143226; AAD44197.1; -.
DR   EMBL; AF143224; AAD44197.1; JOINED.
DR   EMBL; AF143225; AAD44197.1; JOINED.
DR   EMBL; AF143222; AAD44198.1; -.
DR   EMBL; AF141021; AAD42072.1; -.
DR   EMBL; AF152242; AAD46774.1; -.
DR   EMBL; AF152240; AAD46774.1; JOINED.
DR   EMBL; AF152241; AAD46774.1; JOINED.
DR   EMBL; AF146592; AAD46769.1; -.
DR   MIM; 603937; -.
DR   MIM; 180100; -.
DR   InterPro; IPR003533; DCX.
DR   SMART; SM00537; DCX; 2.
DR   PROSITE; PS50309; DC; 2.
KW   Vision; Retinitis pigmentosa; Retinal protein; Polymorphism; Repeat.
FT   DOMAIN       36    118       DOUBLECORTIN 1.
FT   DOMAIN      154    233       DOUBLECORTIN 2.
FT   DOMAIN      268    273       POLY-SER.
FT   DOMAIN      671    675       POLY-LYS.
FT   DOMAIN     1687   1691       POLY-SER.
FT   VARIANT     872    872       R -> H.
FT                                /FTId=VAR_007810.
FT   VARIANT     985    985       N -> Y.
FT                                /FTId=VAR_007811.
FT   VARIANT    1670   1670       A -> T.
FT                                /FTId=VAR_007812.
FT   VARIANT    1691   1691       S -> P.
FT                                /FTId=VAR_007813.
FT   VARIANT    2033   2033       C -> Y.
FT                                /FTId=VAR_007814.
SQ   SEQUENCE   2156 AA;  240659 MW;  55AEDBEC43D6A507 CRC64;
```

```
  Query Match          3.4%;  Score 190.5;  DE 1;  Length 2156;
  Best Local Similarity  20.3%;  Pred. No. 0.028;
  Matches  204;  Conservative  136;  Mismatches  373;  Indels  291;  Gaps   52;

Qy      93 ASSSPFSVAKFRRWDCSKCLDKLKTSDNGTAPRTLPAKQNGTSDGCSITFVRSTFVPASV 152
           |   ||   |   :|  | | |     | : |  | | |  |:   |    || : :|
Db     222 AGREPF---KPGNYDIQKYL--LPARLPGISQRVYP-KGNAKSESRKI----STHMSSSS 271

Qy     153 GSQKVSPSTQSSQGKNA--DRSTLPKS--VQEGNDSKCNAPSGKNGAAEANTDSPMKDLQ 208
             ||  | |:: :   :   | | :|:    | |||: | |             | :|
Db     272 RSQIYSVSSEKTHNNDCYLDYSFVPEKYLALEKNDSQ-NLPI----------YPSED-- 317

Qy     209 GPAQNYDVAANV--SEDNTSVDVGALPEVPQITWHIEVNGADQPPSTPKLSEVVLK---R 263
                 |:  ::  ::| |           :|  ::|    :   | | :  | |
Db     318 ------DIEKSIIFNQDGT----------MTVEMKVRFRIKEEETIKWTTTVSKTGPS 359

Qy     264 NEDEN------GKTEET---LVAEQCNLTKDPNPM---SGKE----------RDQVAEQ 300
           | ||       |:||     |    |: : | :||   | :|           |||||
```

```
Db     360 NNDEKSEMSFPGRTESRSSGLKLAACSFSADVSPMERSSNQEGSLAEEINIQMTDQVAET 419

Qy     301 C------NLTKDPKPVSG---QKCEQICNEPCEEVVLKRSSK---------SKRKTDKKL 342
           |      | | |   : |   |   :    |   :   | |         |: :  :|:
Db     420 CSSASWENATVDTDIIQGTQDQAKHRFYRPPTPGLRRVRQKKSVIGSVTLVSETEVQEKM 479

Qy     343 MKKQQHSKKRTAQADVSDAKLCRRKPKKVR------LLSEIINANQVEDSRSDEVHRENA 396
           : :  :|::| :  : |:  :      |:       :| :| | :|:|:|  :     :
Db     480 IGQFSYSEERESGENKSEYHMFTHSCSKMSSVSNKPVLVQINNNDQMEESSLERKKENSL 539

Qy     397 ADPCEDDRSTIPVPMEVSMDIPVSNHTVGEDGLKSSKNKTKRKYSDVV-----DDGSSLM 451
                   | :  :  ::::  :|      || |: :       |||     |: : :
Db     540 LKSSAISAGVIEITSQKMLEMSHNN------GLPSTISNNSIVEEDVVDCVVLDNKTGIK 593

Qy     452 NWLNGKKKRTGSVHHTVAHPAGNLSNKKVTP-TASTQHDDENDTENGLDTNMHKTDVCQH 510
           |:     |  |:             :| : :| :|   |   |:  :| | |: :    :
Db     594 NF-----KTYGN-----------TNDRFSPISADATHFSSNN--SGTDKNISEAPASEA 634

Qy     511 VSEISTQ--RCSSKGKTAGLSK-----GKTHSAASTKYGGESTRNGQNIHVLSAEDQCQM 563
           | :: :  |  ::    ||:|      |  |: ::|   :| :   |    |      |:
Db     635 SSTVTARIDRLINEFAQCGLTKLPKNEKKILSSVASKKKKKSRQQAIN----SRYQDGQL 690

Qy     564 ETENSVLSHSAKV----------------SPAEHDIQIMSDLHEQSLPKKKKKQKLEVTR 607
           |:  :|: : ::                 || : .|    ||           || :
Db     691 ATK-GILNKNERINTKGRITKEMIVQDSDSPLKGGILCEEDL----------QKSDTVI 738

Qy     608 EKQTMIDDIPMDIVELLAKNQHERQ----------LMTETDCSDINRI-----QSKTTA 651
           |  |       ::   ::|| |  :           |:|:     :|:|     : :
Db     739 ESNTFCS--KSNLNSTISKNFHRNKLNTTQNSKVQGLLTKRKSRSLNKISLGAPKKREIG 796

Qy     652 DDDCVIVAAKDGSDYASSVFDTNS--------QQKSLASQSTQKELQGHLALTTQESPHP 703
             | |     : | |  | |:  |        :||                       |
Db     797 QRDKVF--PHNESKYCKSTFENKSLFHVFNILEQK-----------------------P 830

Qy     704 QNFQSTQEQQTHLRMEEMVTIAASSPLFSHHDDQYIAEAPTEHWGRKDAKKLTWEQFKAT 763
           ::| : | |              || |        :::     |   | ||   :: ||:
Db     831 KDFYAPQSQAE----------VASGYLRGMAKKSLVSKVTDSHITLKSQKKRKGDKVKAS 880

Qy     764 T-RNSPAATCGAQFRPGIQAVDLTSTHVMGSSSNYASRQPVIAPLDRYAERAVNQVHARN 822
             :  || |     ::  |        |  ||   |  :  :: |
Db     881 AILSKQHATTRANSLASLKKPDFPEAIAHHSIQNYI--QSWLQNINPY---------- 926

Qy     823 FPSTIATMEASKLCDRRNAGQVVLYPKESMPATHLLRMMDPSTLASFPNYGTSSRNQMES 882
             |:  :::: :|  ||   ||          |           || |     | |  |||
Db     927 --PTLKPIKSAPVC--RNETSVVNCSNNSFSGN------DPHT-----NSGKISNFVMES 971

Qy     883 QLHNSQYAHNQYKGSTSTSYGSNL--NGKIPLTFEDLSRHQLHDLHRPLRPHPRVGVLGS 940
           | :: |     | |    | ||   |       |     ||:          :||
Db     972 NKHITKIA-----GLT----GDNLCKEGDKSFIANDTGEEDLHETQ----------VGS 1011

Qy     941 LLQKEIANWSENCGTQSGYKLGVSTGITSHQMNRKEHFEALNSG 984
           |   :   |:| | |       : |  |  | | |  |  ||
Db    1012 LNDAYLVPLHEHC-TLS------QSAINDH--NTKSHIAAEKSG 1046
```

```
RESULT  11
MAPB_RAT
ID   MAPB_RAT       STANDARD;      PRT;  2459 AA.
AC   P15205; Q62958; Q9ER21; Q9QW92;
DT   01-APR-1990 (Rel. 14, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Microtubule-associated protein 1B (MAP 1B) (Neuraxin) [Contains: MAP1
DE   light chain LC1].
GN   MAP1B.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE OF 1-142 FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=Testis;
RX   MEDLINE=96257242; PubMed=8666295;
```

```
RA   Liu D., Fischer I.;
RT   "Isolation and sequencing of the 5' end of the rat microtubule-
RT   associated protein (MAP1B)-encoding cDNA.";
RL   Gene 172:307-308(1996).
RN   [2]
RP   SEQUENCE OF 96-2459 FROM N.A., DOMAIN, AND INDUCTION.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=Brain, and Glial tumor;
RX   MEDLINE=92347374; PubMed=1639092;
RA   Zauner W., Kratz J., Staunton J., Feick P., Wiche G.;
RT   "Identification of two distinct microtubule binding domains on
RT   recombinant rat MAP 1B.";
RL   Eur. J. Cell Biol. 57:66-74(1992).
RN   [3]
RP   SEQUENCE OF 1541-2459 FROM N.A., AND TISSUE SPECIFICITY.
RC   TISSUE=Spinal cord;
RX   MEDLINE=90059871; PubMed=2555150;
RA   Rienitz A., Grenningloh G., Hermans-Borgmeyer I., Kirsch J.,
RA   Littauer U.Z., Prior P., Gundelfinger E.D., Schmitt B., Betz H.;
RT   "Neuraxin, a novel putative structural protein of the rat central
RT   nervous system that is immunologically related to microtubule-
RT   associated protein 5.";
RL   EMBO J. 8:2879-2888(1989).
RN   [4]
RP   DEVELOPMENTAL STAGE, AND PHOSPHORYLATION.
RX   MEDLINE=97405699; PubMed=9260743;
RA   Ma D., Nothias F., Boyne L.J., Fischer I.;
RT   "Differential regulation of microtubule-associated protein 1B (MAP1B)
RT   in rat CNS and PNS during development.";
RL   J. Neurosci. Res. 49:319-332(1997).
CC   -!- FUNCTION: The function of brain MAPS is essentially unknown.
CC       Phosphorylated MAP1B may play a role in the cytoskeletal changes
CC       that accompany neurite extension. Possibly MAP1B Binds to at least
CC       two tubulin subunits in the polymer, and this bridging of subunits
CC       might be involved in nucleating microtubule polymerization and in
CC       stabilizing microtubules.
CC   -!- SUBUNIT: 3 different light chains, LC1, LC2 and LC3, can associate
CC       with MAP1A and MAP1B proteins.
CC   -!- TISSUE SPECIFICITY: Nervous system (spinal cord, brain stem,
CC       cerebellum and cerebrum). Not expressed in liver, spleen, kidney,
CC       heart or muscle.
CC   -!- DEVELOPMENTAL STAGE: In cerebral cortex, spinal cord and sciatic
CC       nerve levels are high early in development but decrease during
CC       postnatal development and are low in adults. In dorsal root
CC       ganglia levels remain high throughout development.
CC   -!- INDUCTION: By nerve growth factor.
CC   -!- DOMAIN: Has a highly basic region with many copies of the sequence
CC       KKEE and KKEI/V, repeated but not at fixed intervals, which is
CC       responsible for the binding of MAP1B to microtubules.
CC   -!- PTM: LC1 is coexpressed with MAP1B. It is a polypeptide generated
CC       from MAP1B by proteolytic processing. It is free to associate with
CC       both MAP1A and MAP1B. It interacts with the amino-terminal region
CC       of MAP1B (By similarity).
CC   -!- PTM: Phosphorylated.
CC   -!- SIMILARITY: TO MAP1A.
CC   -!- CAUTION: A C-terminal fragment of this protein (residues 1597 to
CC       2459) was originally described as neuraxin in ref.3.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U52950; AAB17068.1; -.
DR   EMBL; X60370; CAC16162.1; -.
DR   EMBL; X16623; CAA34620.1; ALT_SEQ.
DR   PIR; S06017; S06017.
DR   InterPro; IPR000102; MAP1B_neuraxin.
DR   Pfam; PF00414; MAP1B_neuraxin; 10.
DR   PROSITE; PS00230; MAP1B_NEURAXIN; 8.
KW   Microtubules; Repeat; Phosphorylation.
```

```
FT   CHAIN        ?   2459       MAP1 LIGHT CHAIN LC1.
FT   REPEAT    1869   1885       MAP1B 1.
FT   REPEAT    1886   1902       MAP1B 2.
FT   REPEAT    1903   1919       MAP1B 3.
FT   REPEAT    1920   1936       MAP1B 4.
FT   REPEAT    1937   1953       MAP1B 5.
FT   REPEAT    1954   1970       MAP1B 6.
FT   REPEAT    1988   2004       MAP1B 7.
FT   REPEAT    2005   2021       MAP1B 8.
FT   REPEAT    2022   2038       MAP1B 9.
FT   REPEAT    2039   2055       MAP1B 10.
FT   DOMAIN     559   1035       GLU-RICH.
FT   DOMAIN     588    786       LYS-RICH (HIGHLY BASIC, CONTAINS MANY
FT                               KKEE AND KKEI/V REPEATS).
FT   DOMAIN    2224   2312       LYS-RICH.
FT   CONFLICT   127    127       M -> V (IN REF. 1).
FT   CONFLICT   140    140       T -> S (IN REF. 1).
FT   CONFLICT  2112   2112       R -> K (IN REF. 3).
FT   CONFLICT  2169   2169       L -> I (IN REF. 3).
SQ   SEQUENCE  2459 AA;  269497 MW;  2E3F6872DEDB8BA2 CRC64;
```

```
  Query Match             3.3%;  Score 184;  DB 1;  Length 2459;
  Best Local Similarity   18.7%;  Pred. No. 0.072;
  Matches  195;  Conservative  166;  Mismatches  421;  Indels  260;  Gaps   43;

Qy    167 KNADRSTLPKSVQEGNDSKCNAPSGKNGAAEANTDS--PMKDLQGPAQNYDVAANVSEDN 224
          |: |    | : |:    : : |  |    :   ||   :|    |  :  |     |
Db    505 KHLDFLKQPLATQKDLTGQVSTPPVKQVKLKQRADSRESLKPATKPLSSKSVRKESKE-- 562

Qy    225 TSVDVGALPEVPQITWHIEVNGADQPPSTPKL---SEVVLKRNEDENGKTEE--TLVAEQ 279
                  | |: |       | |   |||:    :|::|:  |: || |   ::  ::
Db    563 ---------EAPEAT------KASQVEKTPKVESKEKVIVKK--DKPGKVESKPSVTEKE 605

Qy    280 CNLTKDPNPMSGKERDQVAEQCNLTKDPKPVSGQKC--EQICNEPCEEVVLKRSSKSKRK 337
            :: :|:    : :|||:      || |:  |   ::|  :| |:   |   :  :|
Db    606 VPSKEEQSPV----KAEVAEKAATESKPK-VTKDKVVKKEIKTKPEEKKEEKPKKEVAKK 660

Qy    338 TDKKLMKKQQHSKKRTAQADVSDAKLCRRKPKKVRLLSEIINANQVEDSRSDEVHRENAA 397
           ||  :|| :  ||  |: ::   |   :| :|  |  |:     ::|::  || ::
Db    661 EDKTPLKKDEKPKKEEAKKEI---KKEIKKEEKKELKKEVKKETPLKDAKK-EVKKDEKK 716

Qy    398 DPCEDDRSTIPVPMEVSMDI----PVSN---------------HTVGEDGLKSSKNKTKR 438
          :  ::::       ::| ||    |:|:                   :: : : | | |
Db    717 EVKKEEKEPKKEIKKISKDIKKSTPLSDTKKPAALKPKVAKKEEPTKKEPIAAGKLKDKG 776

Qy    439 KYSDVVDDGSSLMNWLNGKKKRTGSVHHTVAHPAGNLSN---KKVTPTASTQHDDENDTE 495
          |   :  :| :       :   |      ||  ||  ::   |::    |     |: |:
Db    777 KVKVIKKEGKTT------EAAATAVGTAAVAAAAGVAASGPAKELEAERSLMSSPEDLTK 830

Qy    496 NGLDTNMHKTDVCQHV-------------SEISTQRCSSKGKTAGLSKGKTHS----AAS 538
          :  :    : || : :                |          |   :|||   |     :
Db    831 DFEELKAEEIDVAKDIKPQLELIEDEEKLKETEPGEAYVIQKETEVSKGSAESPDEGITT 890

Qy    539 TKYGGESTRNGQNIHVLSAE--------------------------------DQCQMET 565
          |:  ||  :  : :  :  :                                  :: : :
Db    891 TEGEGECEQTPEELEPVEKQGVDDIEKFEDEGAGFEESSEAGDYEEXAETEEAEEPEEDG 950

Qy    566 ENSVLSHSAKVSPAEHDIQIMSDLHEQSLPKKKKKQKLEVTREKQTMIDD-IPMDIVELL 624
          |::|   ::| || |          :: : | :    ::  ||     ||    |: | |
Db    951 EDNVSGSASKHSPTE----------DEEIAKAEADVHIKEKRESVASGDDRAEEDMDEAL 1000

Qy    625 AK---NQHERQLMTETDCSDINRIQ----SKTTADDDCVIVAAK--------DGSDY--- 666
            |    | | :   | | ::  | :     || |:|  : |  |         |  |:
Db   1001 EKGEAEQSEEEGEEEEDKAEDAREEDHEPDKTEAEDYVMAVVDKAAEAGVTEDQYDFLGT 1060

Qy    667 --------------ASSVFDT----NSQQKSLASQSTQKELQ--------GHLALTTQES 700
                        |||: |      |: :: ||    :| |         |:   | : |
Db   1061 PAKQPGVQSPSREPASSIHDETLPGGSESEATASDEENREDQPEEFTATSGYTQSTIEIS 1120

Qy    701 --PHPQNFQS---------TQEQQTHLRMEEMVTIAA-SSPLFSHHDDQYIAEAPTEHWG 748
            | | :  |           | ::|    :| | |    | |:|    :|         |
Db   1121 SEPTPMDEMSTPRDVMTDETNNEETESPSQEFVNITKYESSLYS---QEYSKPVVASFNG 1177
```

```
Qy     749 RKDAKKLTWEQFKATTRNSPAATCGA----------------QFRP---------GIQAV 783
Db    1178 LSDGSKT--DATDGRDYNASASTISPPSSMEEDKFSKSALRDAYRPEETDVKTGAELDIK 1235

Qy     784 DLTSTHVMGSSSNYASRQPVIAPLDR--YAERAVNQVHARNFPSTIATMEASKLCDRRNA 841
Db    1236 DVSDERLSPAKSPSLSPSPP-SPIEKTPLGERSVNFSLTPNEIKASAEGEATAVVS-PGV 1293

Qy     842 GQVVLYPKESMPATHLLRMMDPST----LASFPNYGTSSRNQMESQL-----HNSQ---- 888
Db    1294 TQAVVEEHCASPEEKTLEVVSPSQSVTGSAGHTPYYQSPTDEKSSHLPTEVTENAQAVPV 1353

Qy     889 -YAHNQYKGSTSTSYGSNLNGKIPLTFEDLSRHQLHDLHRPLRPHPRVGVLG------SL 941
Db    1354 SFEFTEAKDENERSSISPMDEPVP-----DSESPIEKVLSPLRSPPLIGSESAYEDFLSA 1408

Qy     942 LQKEIANWSENCGTQSGYKLGVSTGITSHQMNRKEHFHALNSGMFSAKWNALQLGSV--- 998
Db    1409 DDKALGRRSESPFEGKNGKQGFS--------DKESPVSDLTSDLYQDKQEEKRAGFIPIK 1460

Qy     999 --------SSSADFLSARNSIA 1012
Db    1461 EDFSPEKKASDAEIMSSQSALA 1482


RESULT 12
ZIP1_YEAST
ID   ZIP1_YEAST     STANDARD;      PRT;   875 AA.
AC   P31111;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Synaptonemal complex protein ZIP1.
GN   ZIP1 OR YDR285W OR D9819.9.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BR1824-3B;
RX   MEDLINE=93161412; PubMed=7916652;
RA   Sym M., Engebrecht J.A., Roeder G.S.;
RT   "ZIP1 is a synaptonemal complex protein required for meiotic
RT   chromosome synapsis.";
RL   Cell 72:365-378(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA   Favello A., Fulton L., Gattung S., Greco T., Kirsten J.,
RA   Kucaba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,
RA   Johnson D., Johnston L., Langston Y., Latreille P., Le T.,
RA   Mardis E., Menezes S., Miller N., Nhan M., Pauley A., Peluso D.,
RA   Rifken L., Riles L., Taich A., Trevaskis E., Vignati D.,
RA   Wilcox L., Wohldman P., Vaudin M., Wilson R., Waterston R.;
RL   Submitted (MAR-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: REQUIRED FOR MEIOTIC CHROMOSOME SYNAPSIS AND CELL CYCLE
CC       PROGRESSION. MAY ACT AS A MOLECULAR ZIPPER TO BRING HOMOLOGOUS
CC       CHROMOSOMES IN CLOSE APPOSITION. ZIP1 MAY ENCODE THE TRANSVERSE
CC       FILAMENTS OF THE SYNAPTONEMAL COMPLEX.
CC   -!- SUBCELLULAR LOCATION: SYNAPSED MEIOTIC CHROMOSOMES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L06487; AAA35239.1; -.
DR   EMBL; U51031; AAB64474.1; -.
DR   PIR; A45173; A45173.
DR   PIR; S30868; S30868.
DR   SGD; S0002693; ZIP1.
KW   Nuclear protein; Meiosis; Coiled coil.
FT   DOMAIN      177    333       COILED COIL (POTENTIAL).
FT   DOMAIN      397    438       COILED COIL (POTENTIAL).
FT   DOMAIN      456    752       COILED COIL (POTENTIAL).
FT   CONFLICT     55     55       T -> A (IN REF. 1).
SQ   SEQUENCE   875 AA;  100035 MW;  674F12625CD9DDFD CRC64;


  Query Match             3.3%;  Score 183;  DB 1;  Length 875;
  Best Local Similarity   19.3%;  Pred. No. 0.022;
  Matches  161;  Conservative  149;  Mismatches  362;  Indels  162;  Gaps   34;

Qy     129 AKQNGTSDGCSITFVRSTFVPASVGSQKVSPSTQSSQGKNADRSTLPKSVQEGNDSKCNA 188
Db     109 AIENDTDEDFEITEVREV-------SEGVAKETKESHGDPNDSETTLK------DSKMHE 155

Qy     189 PSGKNGAA-------EANTDSPMKDLQGPAQNYDVAANVSEDNTSVDVGALPEVPQITWH 241
Db     156 YTMTNGKAPLHTSINNSSTSSNDVLLEAFTNTQRICSNLKQE--------LQKQQQDNAK 207

Qy     242 IEVNGADQPPSTPKLSEVVLKRNEDENGKTE---ETLVAEQCNLTKDPNPMSGKERDQVA 298
Db     208 LKVRLQSYASNSDKINEKV--------GKYKSCLETLQERIATLTSHKNNQETKLKDL-- 257

Qy     299 EQCNLTKDPKPVSGQKCE-QICNEPCEEVVLKRSSKSKRKTDKKLMKKQQHSKKRTAQAD 357
Db     258 -RQNHQLYQRRISGFKTSIENLNKTINDL-----GKNKKEADAELMKKGKEIEYLKRELD 311

Qy     358 VSDAKLCRRKPKKVRLLSEIINANQVEDSRS-------DEVHRENAADPCEDDRSTIPVP 410
Db     312 DCSGQLSEEKIKNSSLIQE-MGKNREEMIKSIENFFSEDKAHH---------------- 353

Qy     411 MEVSMDIPVSNHTVGEDGLKSSKNKTKRKYSDVVDDGSSLMNWLNGKKKRTGSVHHTVAH 470
Db     354 --------LLQFNKFEERVHDLFEKKLQKHFDVAKD------TLN-----VGLRNTTV-- 392

Qy     471 PAGNLSNKKVTPTASTQHDDENDTENGLDTNMHKTDVCQHVSEISTQRCSSKGKTAGLSK 530
Db     393 ---ELSSNTETMLKQQYEDIKENLEQKMSSS--KDEMAKTINELSV---TQKGLIMGVQE 444

Qy     531 GKTHSAASTKYGGESTRNGQNIHVLSAEDQC---QMETENSVLSHSAK-VSPAEHDIQIM 586
Db     445 ELLTSSGNIQTALVSEMNNTRQELLDDASQTAKNYASLENLVKAYKAEIVQSNEYEERIK 504

Qy     587 SDLHEQSLPKKKKKQKLEVTREKQTMIDDIPMDIVELLAKN------QHERQLMTETD-- 638
Db     505 HLESERSTLSSQKNQIISSLGTKEAQYEDL---VKKLEAKNIEISQISGKEQSLTEKNEN 561

Qy     639 -CSDINRIQSKTTADDDCVIVAAKDGSDYASSVFDTNSQQKSLASQ--STQKELQGHLAL 695
Db     562 LSNELKKVQDQLEKLNNLNITTK---SNYENKISSQNEIVKALVSENDTLKQRIQQLVEI 618

Qy     696 TTQE----SPHPQNFQSTQEQQTHLRMEEMVTIAASSPLFSHHDDQYI------AEAPTE 745
Db     619 KENEQKDHTTKLEAFQKNNEQLQKLNVE--VVQLKAHELELEEQNRHLKNCLEKKETGVE 676

Qy     746 HWGRKDAKKLTWEQFKATTRNSPAATCGAQFRPGIQAVDLTSTHVMGSSSNYASRQPVIA 805
Db     677 E-SLSDVKTLKQQVIVLKSEKQDITAEKLELQDNLESLEEVTKNL----------QQKVQ 725

Qy     806 PLDRYAERAVNQV-----HARNFPSTIATMEASKLCD--RRNAGQVVLYPKESMPATHLL 858
Db     726 SQKRELEQKIKELEEIKNHKRNEPSKKGTQNFTKPSDSPKKNATTSNLFPNNS-AAIH-- 782

Qy     859 RMMDPSTLASFPNYGTSSRNQMESQLHNSQYAHNQYKGSTSTSYGSNLNGKIPL 912
Db     783 -----SPMKKCPKVDHISKSRINSSKETSKF-NDEFDLSSSSNDDLELTNPSPI 830
```

```
RESULT  13
MAPB_MOUSE
ID   MAPB_MOUSE     STANDARD;      PRT;  2464 AA.
AC   P14873;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Microtubule-associated protein 1B (MAP 1B) (MAP1.2) (MAP1(X))
DE   [Contains: MAP1 light chain LC1].
GN   MAP1B OR MTAP1B OR MTAP5.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., AND DOMAIN.
RC   STRAIN=SWISS WEBSTER; TISSUE=Brain;
RX   MEDLINE=90094539; PubMed=2480963;
RA   Noble M., Lewis S.A., Cowan N.J.;
RT   "The microtubule binding domain of microtubule-associated protein
RT   MAP1B contains a repeated sequence motif unrelated to that of MAP2
RT   and tau.";
RL   J. Cell Biol. 109:3367-3376(1989).
CC   -!- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN.
CC       PHOSPHORYLATED MAP1B MAY PLAY A ROLE IN THE CYTOSKELETAL CHANGES
CC       THAT ACCOMPANY NEURITE EXTENSION. POSSIBLY MAP1B BINDS TO AT LEAST
CC       TWO TUBULIN SUBUNITS IN THE POLYMER, AND THIS BRIDGING OF SUBUNITS
CC       MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN
CC       STABILIZING MICROTUBULES.
CC   -!- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE
CC       WITH MAP1A AND MAP1B PROTEINS.
CC   -!- DOMAIN: Has a highly basic region with many copies of the sequence
CC       KKEE and KKEI/V, repeated but not at fixed intervals, which is
CC       responsible for the binding of MAP1B to microtubules.
CC   -!- PTM: LC1 IS COEXPRESSED WITH MAP1B. IT IS A POLYPEPTIDE GENERATED
CC       FROM MAP1B BY PROTEOLYTIC PROCESSING. IT IS FREE TO ASSOCIATE WITH
CC       BOTH MAP1A AND MAP1B. IT INTERACTS WITH THE AMINO-TERMINAL REGION
CC       OF MAP1B.
CC   -!- SIMILARITY: TO MAP1A.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X51396; CAA35761.1; -.
DR   PIR; S07549; QRMSP1.
DR   MGD; MGI:1306778; Mtap1b.
DR   InterPro; IPR000102; MAP1B_neuraxin.
DR   Pfam; PF00414; MAP1B_neuraxin; 10.
DR   PROSITE; PS00230; MAP1B_NEURAXIN; 7.
KW   Microtubules; Repeat; Phosphorylation.
FT   CHAIN         ?   2464       MAP1 LIGHT CHAIN LC1.
FT   REPEAT     1874   1890       MAP1B 1.
FT   REPEAT     1891   1907       MAP1B 2.
FT   REPEAT     1908   1924       MAP1B 3.
FT   REPEAT     1925   1941       MAP1B 4.
FT   REPEAT     1942   1958       MAP1B 5.
FT   REPEAT     1959   1975       MAP1B 6.
FT   REPEAT     1993   2009       MAP1B 7.
FT   REPEAT     2010   2026       MAP1B 8.
FT   REPEAT     2027   2043       MAP1B 9.
FT   REPEAT     2044   2060       MAP1B 10.
FT   DOMAIN      589    787       LYS-RICH (HIGHLY BASIC, CONTAINS MANY
FT                                KKEE AND KKEI/V REPEATS).
SQ   SEQUENCE   2464 AA;  270408 MW;  FBD3DD99CFDBDA87 CRC64;

  Query Match             3.3%;  Score 182.5;  DB 1;  Length 2464;
  Best Local Similarity   18.9%;  Pred. No. 0.086;

  Matches  175;  Conservative  142;  Mismatches  357;  Indels  251;  Gaps   41;

Qy    233 PEVPQITWHIEVNGAD--QPPSTPKLSEVVLKRNEDENGKTEETLVAEQCN--------L 282
          | | |:      : :  :| : |  |: | | :::|  :  :|   |:         |
Db    528 PPVKQVKLKQRADSRESLKPATKPVASKSVRKESKEETPEVTKTSQVEKTPKVESKEKVL 587

Qy    283 TKDPNPMSGKERDQVAE-QCNLTKDPKPVSGQKCEQICNEPCEEVVLKRSSKSKRKTDK- 340
           |   |:  : :  | | : :  ::  ||              ||  |::::|| |  |
Db    588 VKKDKPVKTESKPSVTEKEVSSKEEQSPVKA-----------EVAEKQATESKPKVTKD 635

Qy    341 KLMKKQQHSKKRTAQADVSDAKLCRRKPKKVRLLSEIIN------ANQVEDSRSDEVHRE 394
          |::||:  :|    : :          ||||     |::          : |  | :|| :|
Db    636 KVVKKEIKTKLEEKKEE---------KPKK-----EVVKKEDKTPLKKDEKPRKEEVKKE 681

Qy    395 NAADPCEDDRSTIPVPMEVSMDIPVSNHTVGEDGLKSSKNKTKRKYSDVVDDGSSLMNWL 454
             :  :::|  :    ||  : |          || :| : |:      ::   :
Db    682 IKKEIKKEERKEL--KKEVKKETP----------LKDAKKEVKK------EEKKEVKKEE 723

Qy    455 NGKKKRTGSVHHTVAHPAGNLSNKKVTPTASTQHDDENDTENGLDTNMHKTDVCQHVSEI 514
             ||    :   :          || || : |:        : |   : | :        :
Db    724 KEPKKEIKKISKDI---------KKSTPQSDTK------KPSALKPKVAKKEESTKKEPL 768

Qy    515 STQRCSSKGKTAGLSK-GKTHSAASTKYGGESTRNGQNIHVLSAEDQC------QMETEN 567
          :  :   |||   : | |||  ||:|  |  :|      |:::|           ::| |
Db    769 AAGKLKDKGKVKVIKKEGKTTEAAATAVGTAAT----TAAVVAAAGIAASGPVKELEAER 824

Qy    568 SVLSHSAKVSPAE----------HDIQIMSDLHEQSLPKKKKKQKLEVTREKQTMIDDIP 617
          |::|     || :            :| :  |:  | |   : ::||: |:  :  :
Db    825 SLMS-----SPEDLTKDFEELKAEEIDVAKDIKPQ-LELIEDEEKLKETQPGEAYVIQKE 878

Qy    618 MDIVELLAKNQHERQLMT--ETDC----SDINRIQSKTTADDDCVIVAAKDGSDYASSVF 671
          :: :  |::  |    |  | :|     ::  :: |   ||  :     :|: :  |
Db    879 TEVSKGSAESPDEGITTTEGEGECEQTPEELEPVE-KQGVDD--IEKFEDEGAGFEESSE 935

Qy    672 DTNSQQKSLASQSTQKELQGH---LALTTQESPHPQNFQSTQEQQTHLRMEEMVTIAASS 728
             : ::|:   :: : |  |        :: ||   :  :  |   ||: :    ::
Db    936 TGDYEEKAETEEAEEPEEDGEDNASGSASKHSPTEDDESAKAEADVHLKEKRESVVSGDD 995

Qy    729 PLFSHHDD-------QYIAEAPTEHWGRKDAKKLTWEQFKATTRN------SPAATCGA- 774
                ||        :   |   |    :||::  :|  |    :          ||  |
Db    996 RAEEDMDDVLEKGEAEQSEEEGEEEDKAEDAREEGYEPDKTEAEDYVMAVADKAAEAGVT 1055

Qy    775 --QF--------RPGIQAVDLTSTHVMGSSSNYASRQPVIAPLDRYAERAVNQVHARNFP 824
            |:        :||||:                ||:|             : :|    |
Db   1056 EEQYGYLGTSAKQPGIQS--------------PSREP-----------ASSIHDETLP 1088

Qy    825 STIATMEASKLCDRRNAGQVVLYPKE--------------SMPATHLLRMMDPSTLAS-- 868
             :  ||:   :      |    |:|                |   | :  |  |  : |
Db   1089 GG-SESEATASDEENREDQ----PEEFTATSGYTQSTIEISSEPTPMDEMSTPRDVMSDE 1143

Qy    869 FPNYGTSSRNQ-------MESQLHNSQYAH--------------------NQYKGSTST- 900
            |  | | :|         || |:: :|:                       |  | ||
Db   1144 TNNEETESPSQEFVNITKYESSLYSQEYSKPAVASFNGLSEGSKTDATDGKDYNASASTI 1203

Qy    901 ----------------------SYGSNLNGKIPLTFEDLSRHQLHDLHRP-LRPHPRVGVL 938
                               |    |      |  :|:|  :|     | | | |   :
Db   1204 SPPSSMEEDKFSKSALRDAYCSEEKELKASAELDIKDVSDERLSPAKSPSLSPSPPSPIE 1263

Qy    939 GSLLQKEIANWS----ENCGTQSGYKLGVSTGITSHQMNRKEHFEALNSGMFSAKWNALQ 994
             : | :   |:|    |   :  |    || |:|  |   :||       | :   |:
Db   1264 KTPLGERSVNFSLTPNEIKVSAEGEARSVSPGVT--QAVVEEH-------CASPEEKTLE 1314

Qy    995 LGSVSSSADFLSARNSIAQSWTRGK 1019
          : | | |    :      || |  |
Db   1315 VVSPSQSVTGSAGHTPYYQSPTDEK 1339


RESULT  14
P531_HUMAN
ID   P531_HUMAN     STANDARD;      PRT;  1972 AA.
AC   Q12888;
DT   15-JUL-1998 (Rel. 36, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
```

```
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Tumor suppressor p53-binding protein 1 (p53-binding protein 1)
DE   (53BP1).
GN   TP53BP1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND SUBCELLULAR LOCATION.
RC   TISSUE=Skeletal muscle;
RX   MEDLINE=98421532; PubMed=9748285;
RA   Iwabuchi K., Li B., Massa H.F., Trask B.J., Date T., Fields S.;
RT   "Stimulation of p53-mediated transcriptional activation by the
RT   p53-binding proteins, 53BP1 and 53BP2.";
RL   J. Biol. Chem. 273:26061-26068(1998).
RN   [2]
RP   SEQUENCE OF 946-1972 FROM N.A.
RX   MEDLINE=94286584; PubMed=8016121;
RA   Iwabuchi K., Bartel P.L., Li B., Marraccino R., Fields S.;
RT   "Two cellular proteins that bind to wild-type but not mutant p53.";
RL   Proc. Natl. Acad. Sci. U.S.A. 91:6098-6102(1994).
CC   -!- FUNCTION: ENHANCES P53-MEDIATED TRANSCRIPTIONAL ACTIVATION.
CC   -!- SUBUNIT: BINDS TO THE CENTRAL DOMAIN OF P53.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR. BOTH NUCLEAR AND CYTOPLASMIC
CC       IN SOME CELLS.
CC   -!- SIMILARITY: CONTAINS 2 BRCT DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF078776; AAC62018.1; -.
DR   EMBL; U09477; AAA21596.1; -.
DR   MIM; 605230; -.
DR   InterPro; IPR001357; BRCT.
DR   Pfam; PF00533; BRCT; 4.
DR   SMART; SM00292; BRCT; 2.
DR   PROSITE; PS50172; BRCT; 2.
KW   Nuclear protein; Transcription regulation; Activator; Repeat.
FT   DOMAIN      1724   1848       BRCT 1.
FT   DOMAIN      1864   1964       BRCT 2.
FT   DOMAIN      1642   1646       POLY-SER.
FT   DOMAIN      1760   1764       POLY-GLU.
SQ   SEQUENCE   1972 AA;  213573 MW;  13E2CC8A265F9D2A CRC64;
```

```
  Query Match             3.3%;  Score 181.5;  DB 1;  Length 1972;
  Best Local Similarity   18.6%;  Pred. No. 0.073;
  Matches  180;  Conservative  135;  Mismatches  412;  Indels  243;  Gaps   40;

Qy      30 PVLELTATPRQDAAAEAGVDEPAQHQCEHFSIRGYVALLQKKDPKFC-SLSRIFHDQKKC 88
           |||:: : | | |  | |      :  ||         :   :   | |:|::    ::
Db      57 PVLDVVSNPEQTAGEERGDGNSGFN--EHLKENKVADPVDSSNLDTCGSISQVI---EQL 111

Qy      89 DEHKASSSPFSVAKFRRWDCSKCLDKLKTSDNGTAPRTLPAKQNGTSDGCSITFVRSTFV 148
             :   :||   ::     : :  ::: |  :        |  |:    :  |      |
Db     112 PQPNRTSSVLGMSV----ESAPAVEEEKGEE-------LPQKEKEKEEDTSGNTTHSLGA 160

Qy     149 PASVGSQKVSPSTQSSQGKNADRSTLPKSVQEGNDSKCNAPSGKNGAAEANTDSPMKDLQ 208
             :  ||      : || :: : :|:|  | :          ||    :: : :: :|  :
Db     161 EDTASSQLGFGVLELSQSQDVEENTVPYEVDKEQLQSVTTNSGYTRLSDVDANTAIKHEE 220

Qy     209 GPAQNYDVAANVSEDNTSVDVGALPEVPQITWHIEVNGADQPPS--------TPKLSEVV 260
             ::  :|  |:|        :|  | :  : |      ||          :|| |
Db     221 QSNEDIPIAEQSSKD--------IPVTAQPSKDVHVVKEQNPPPARSEDMPFSPKASVAA 272

Qy     261 LKRNEDENGKTEETLVAEQCNLTKDPNPMSGKERDQVAEQCNLTKDPKPVSGQKCEQICN 320
           ::  |     : : |:     : | | |     :: : :| |     | ||   |
```

```
Db     273 MEAKEQ---LSAQELMESGLQIQKSPEPEVLSTQEDLFDQSN-----KTVSSDGCS---- 320

Qy     321 EPCEEVVLKRSSKSKRKTDKKLMKKQQHSKKRTAQADVSDAKLCRRKPKKVRLLSEIINA 380
                                                 |  :    |     : ||
Db     321 --------------------------------TPSREEGGCSLASTPATTLHLLQLSGQR 348

Qy     381 NQVEDSRSDEVHRENAAD---PCEDD-RST------IPVPMEVSMDIPVSNHTVGEDGLK 430
           : |:|| |      |::|   |  |  |||       |   |   | |:    : |:| :
Db     349 SLVQDSLS-----TNSSDLVAPSPDAFRSTPFIVPSSPTEQEGRQDKPMDTSVLSEEGGE 403

Qy     431 SSKNKTKRKYSDVVDDGSSLMNWLNGKKKRTGSVHHTVAHPAGNLSNKKVTPTASTQHD- 489
             : | :      :::   |       :  |      |   | |:|    :       ||
Db     404 PFQKKLQSGEPVELENPPLLPESTVSPQASTPISQSTPVFPPGSL---PIPSQPQFSHDI 460

Qy     490 -------DENDTENGLDTNMHKTDV---CQHVSEISTQR------CSSKGKTAGLSKGKT 533
                  :|   :   | :|| : :   |   |||  :         |  | :  |    :
Db     461 FIPSPSLEEQSNDGKKDGDMHSSSLTVECSKTSEIEPKNSPEDLGLSLTGDSCKLMLSTS 520

Qy     534 HSAASTKYGGEST----RNGQNIHVLSAEDQCQMETENSVLSHSAKVSPAEHDIQIMSDL 589
             : | |    |:      :|:|  :   ||     |  : ||  ::|  |||:|  :|:
Db     521 EYSQSPKMESLSSHRIDEDGENTQI---ED---TEPMSPVL--NSKFVPAENDSILMN-- 570

Qy     590 HEQSLPKKKKKQKLEVTREKQTMIDDIPMDIVELLAKNQHERQLMTETDCSDINRIQSKT 649
               | :  : :|    :|    |    |  : :||           | |      : :|
Db     571 -----PAQDGEVQLSQNDDKTKGDDTDTRDDISILA----------TGCKG----REET 610

Qy     650 TADDDCVIVAAKDGSDYASSVFDTNSQQKSLASQSTQKEL------QGHLALTTQESPH- 702
             |:| |: :    ||    |    :    |:  |    |:       :|      :| |
Db     611 VAEDVCIDLTCDSGSQAVPSPATRSEALSSVLDQEEAMEIKEHHPEEGSSGSEVEEIPET 670

Qy     703 ----------PQNFQS---------TQEQQTHLRME-----------EMVTIAASSPLF 731
                      :| :|          || |   |: |           |   |:  ||
Db     671 PCESQGEELKEENMESVPLHLSLTETQSQGLCLQKEMPKKECSEAMEVETSVISIDSP-- 728

Qy     732 SHHDDQYIAEAPTEHWGRKDAKKLTWEQFKATTRNSPAATCGAQF---RPGIQAVDLTST 788
               | :   ||      ::  ||:  ||: :|       :    |  :    |
Db     729 ---QKLAILDQELEH-----KEQEAWEE--ATSEDSSVVIVDVKEPSPRVDVSCEPLEGV 778

Qy     789 HVMGSSSNYASRQPVIAP-----LDRYAERAVNQVHARNFPSTIATMEASKLCDRRNAGQ 843
               | ::   | | |      ||   |::|  :  :  |  |  |
Db     779 EKCSDSQSWEDIAPEIEPCAENRLDTKEEKSVE--YEGDLKSGTAETE----------- 824

Qy     844 VVLYPKE---SMPATHLLRMMDPSTLASFPNYGTSSRNQMESQLHNSQYAHNQYKGSTST 900
             | |   | |:  |:|  ||                  ::: :|   |    | | | |
Db     825 ----PVEQDSSQPSLPLVRADDPL-------------RLDQELQQPQ---TQEKTSNSL 863

Qy     901 SYGSNL-NGK 909
           :  | : | |
Db     864 TEDSKMANAK 873
```

```
RESULT  15
CENE_HUMAN
ID   CENE_HUMAN     STANDARD;      PRT;  2663 AA.
AC   Q02224;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Centromeric protein E (CENP-E protein).
GN   CENPE.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93024922; PubMed=1406971;
RA   Yen T.J., Li G., Schaar B.T., Szilak I., Cleveland D.W.;
RT   "CENP-E is a putative kinetochore motor that accumulates just before
RT   mitosis.";
RL   Nature 359:536-539(1992).
RN   [2]
RP   CHARACTERIZATION.
```

```
RX   MEDLINE=95196755; PubMed=7889940;
RA   Thrower D.A., Jordan M.A., Schaar B.T., Yen T.J., Wilson L.;
RT   "Mitotic HeLa cells contain a CENP-E-associated minus end-directed
RT   microtubule motor.";
RL   EMBO J. 14:918-926(1995).
RN   [3]
RP   CHARACTERIZATION.
RX   MEDLINE=98437347; PubMed=9763420;
RA   Chan G.K.T., Schaar B.T., Yen T.J.;
RT   "Characterization of the kinetochore binding domain of CENP-E reveals
RT   interactions with the kinetochore proteins CENP-F and hBUBR1.";
RL   J. Cell Biol. 143:49-63(1998).
CC   -!- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE
CC       KINETOCHORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE
CC       OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT
CC       AND/OR SPINDLE ELONGATION.
CC   -!- SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.
CC   -!- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCHORES DURING
CC       CONGRESSION, RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS
CC       QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.
CC   -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z15005; CAA78727.1; -.
DR   PIR; S28261; S28261.
DR   HSSP; P17119; 3KAR.
DR   MIM; 117143; -.
DR   InterPro; IPR001752; kinesin.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;
KW   Cell cycle; Centromere.
FT   DOMAIN        1    335       KINESIN-MOTOR.
FT   DOMAIN      336   2471       COILED COIL (POTENTIAL).
FT   DOMAIN     2472   2663       GLOBULAR (POTENTIAL).
FT   NP_BIND      86     93       ATP (BY SIMILARITY).
SQ   SEQUENCE   2663 AA;  312087 MW;  CEFC13880C8C8CB8 CRC64;


  Query Match             3.3%;  Score 181.5;  DB 1;  Length 2663;
  Best Local Similarity   18.9%;  Pred. No. 0.11;
  Matches  185;  Conservative  155;  Mismatches  357;  Indels  283;  Gaps   45;

Qy      31 VLELTATPRQDAAAEAGVDEPAQHQCEHFSIRGYV----ALLQKKDPKFCSLSRIFHDQK 86
           : || |   |    :  |:|  :   |   ::  :      | | : :  :|::  |:
Db    1822 IQELKANEHQLITLKKDVNETQKKVSEMEQLKKQIKDQSLTLSKLEIENLNLAQELHE-- 1879

Qy      87 KCDEHKASSSPFSVAK----FRRWDCSKCL--DKLKTSDNGTAPRTLPAKQNGTSDGCSI 140
             :| |      || |     || : :  |  |:|| |   |  | |  :|
Db    1880 NLEEMK------SVMKERDNLRRVEETLKLERDQLKESLQETKARDLEIQQE-------- 1925

Qy     141 TFVRSTFVPASVGSQKVSPSTQSSQGKNADRSTLPKSVQEGNDSKCNAPSGKNGAAEANT 200
                 | : |::   :    : | ::::    :|:  |        |:   :
Db    1926 ------LKTARMLSKEHKETVDKLREKISEKTIQISDIQKDLD------KSKDELQKKIQ 1973

Qy     201 DSPMKDLQGPAQNYDVAANVSEDNTSVDVGALPEVPQITWHIEVN-----GADQPPSTPK 255
           :   |:||           | ||  ::    : |: |:    | |       |    | |
Db    1974 ELQKKELQ--------LLRVKED-VNMSHKKINEMEQLKKQFEPNYLCKCEMDNFQLTKK 2024

Qy     256 LSEV-----VLKRNEDENGKTEETLVAEQCNLTKDPNPMSGKERD--QVAEQCNLTKDPK 308
           | |      :: :  ||  : :|:|  |:         |  ::|   ||  :  |  | :
Db    2025 LHESLEEIRIVAKERDELRRIKESLKMERDQFIATLREMIARDRQNHQVKPEKRLLSDGQ 2084

Qy     309 PVSGQKCEQICNEPCEEV--VLKRSSKSK------------------RKTDKKL----- 342
             |   :   | |  :  :||| |:                       :  |||
Db    2085 ----QHLMESLREKCSRIKELLKRYSEMDDHYECLNRLSLDLEKEIEFHRIMKKLKYVLS 2140

Qy     343 ----MKKQQHSKKRTAQADVSDAKLCRRKPKKVRLLSEIINANQVEDSRSDEVHRENAAD 398
               :|::||      : | |         |:  || :| :  |  |  | |:
Db    2141 YVTKIKEEQHECINKFEMDFID-----EVEKQKELLIKIQHLQQDCDVPSRELR------ 2189

Qy     399 PCEDDRSTIPVPMEVSMDIPVSNHTVGEDGLKSSKNKTKRKYSDVVDDGSS----LMNWL 454
                 : :  :||: :    : :|  :|     | ::  |: :       |  ||
Db    2190 ---------DLKLNQNMDLHIEE--ILKDFSESEFPSIKTEFQQVLSNRKEMTQFLEEWL 2238

Qy     455 ----------NGKKKRTGSVHHTVAHPAGNLSNKKVTP--TASTQHDDENDTENGLDTNM 502
                    || :|    :          |  | ::      ||: :: : | :
Db    2239 NTRFDIEKLKNGIQKENDRICQ-----VNNFFNNRIIAIMNESTEFEERSATIS------ 2287

Qy     503 HKTDVCQHVSEISTQRCSSKGKTAGLSKG----KTHSAASTKYGGESTRNGQNIHVLSAE 558
                    |       | | |   | |      || | ||      :|:: :| || |
Db    2288 ---------KEWEQDLKSLKEKNEKLFKNYQTLKT-SLASGAQVNPTTQDNKNPHVTSRA 2337

Qy     559 DQCQM----------ETENSVLSHSAKVSPAEHDIQIMSDL--------HEQSLPKKKK 599
            |               | : | :   :|:   : ::::: :|:        || :   :|
Db    2338 TQLTTEKIRELENSLHEAKESAMHKESKIIKMQKELEVTNDIIAKLQAKVHESNKCLEKT 2397

Qy     600 KQKLEVTREKQTM--------IDDIPMDIVEL-LAKNQHERQLMTETDCSDINRIQSKTT 650
           |: ::| ::|  :         |:|: | :|:: | | :: ::   |
Db    2398 KETIQVLQDKVALGAKPYKEEIEDLKMKLVKIDLEKMKNAKEFEKE------------- 2443

Qy     651 ADDDCVIVAAKDGSDYASSVFD--TNSQQKSLASQSTQKELQGHLALTTQESPHPQN--- 705
                | | |   :|   |  :  : ::|  :| |          ::    | | |
Db    2444 ------ISATKATVEYQKEVIRLLRENLRRSQQAQDTS-------VISEHTDPQPSNKPL 2490

Qy     706 --------FQSTQE---QQTHLRMEEMVT---------IAASSPLFSHHDDQYIAEAPTE 745
                  |:|:    :  |:|:|: ::         |   : | |  ::|:::    :
Db    2491 TCGGGSGIVQNTKALILKSEHIRLEKEISKLKQQNEQLIKQKNELLS--NNQHLSN-EVK 2547

Qy     746 HWGRKDAKKLTWEQFKATTRNSPAATCGAQFRPGIQAVDLTSTHVMGSSSNYASRQPVIA 805
             |  :  |:   :|   |  |||                 |  | |:    ||::  |
Db    2548 TWKERTLKREAHKQ--VTCENSP----------------KSPKVTGT----ASKKKQIT 2584

Qy     806 PLDRYAERAVNQVHARNFPSTIATMEASKLC--DRRNAGQVVLYPKESMPATHLLRMMDP 863
           |          :|   ||    :   |: | |  | |:          :|:|: | :|  |
Db    2585 P---------SQCKERNLQDPV-PKESPKSCFFDSRS---------KSLPSPHPVRYFDN 2625

Qy     864 STLASFPNYGTSSRNQMESQ 883
           |:|   |    :    ::||
Db    2626 SSLGLCPEVQNAGAESVDSQ 2645


Search completed: September 16, 2002, 22:35:04
Job time: 658 sec
```